LA-UR-19-30247

Approved for public release; distribution is unlimited.

Title: Analysis of Multimodal Wearable Sensor Data to Characterize Social Groups and Influence in Organizations

Author(s): Boero, Riccardo
Hraber, Peter Thomas
Kaufeld, Kimberly Ann
Moore, Elisabeth Ann
Romero-Severson, Ethan
Ambrosiano, John Joseph
Whitton, John Leslie
Sims, Benjamin Hayden

Intended for: Report

Issued: 2019-10-09

Analysis of Multimodal Wearable Sensor Data to Characterize Social Groups and Influence in Organizations

Riccardo Boero, Peter Hraber, Kimberly Kaufeld, Elisabeth Moore, Ethan Romero-Severson, John Ambrosiano, John Whitton, and Benjamin Sims*

Los Alamos National Laboratory

Correspondence*:
PO Box 1663, MS F600, Los Alamos, NM USA 87545
bsims@lanl.gov +1 505 667 5508

SUMMARY

This document reports the results of a 4-month study analyzing data provided to Los Alamos National Laboratory (LANL) by the performer teams for the Multimodal Objective Sensing to Assess Individuals with Context (MOSAIC) project. The MOSAIC performer teams monitored hundreds of volunteers from several organizations in their daily activities, using wearable sensors to capture data about those activities. The project also collected "ground truth" data on individual behavior and psychological state using survey instruments throughout the study. The LANL MOSAIC analysis team approached this data with the goal of capturing and analyzing social dynamics within organizations. We focused on two key analysis objectives: identifying and characterizing groups as the evolve over time, and identifying and modeling patterns of social influence.

The project was conducted in two stages: an initial, intensive "data sprint" at LANL to rapidly identify and test analysis approaches, and a follow-on in-depth study to explore a few analysis approaches in depth. The data sprint was successful in rapidly identifying several promising analysis approaches, and may serve as a model for future analysis efforts at LANL. For the in-depth analysis phase, five of the sprint participants agreed to work together to extend their analyses from the sprint. Through a significant effort to characterize and process the data for analysis, the team identified the Notre Dame performer team's data as the most readily accessible for analysis, and further narrowed its focus to beacon data, which provides a direct proxy for face-to-face interactions. One challenge for interpreting our analysis results was the lack of detailed organizational ground truth, such as information on the size, structure, or work processes engaged in by the organizations represented in the data set.

In characterizing the beacon data, we noted a counter-intuitive property in interactions between beacons. Contact time between two beacons should be registered as the same by both beacons, i.e. beacon A should register the same amount of time in proximity to beacon B as beacon B measures in proximity to beacon A. However, this property does not appear to hold true in the Notre Dame beacon data, in either its raw form or in the processed interaction session data

provided by the Notre Dame team. In order to avoid inconsistencies in our analysis, we chose to force interaction times to be symmetrical by using the maximum interaction time recorded by either member of each pair of participants. Other anomalies suggest possible non-compliance or sensor reliability issues. In addition, the processed interaction session data, which we ended up using for most of our analyses, showed extremely small durations of interaction among study participants. This suggests that this data set may include a large proportion of false negatives, i.e. times when participants were interacting but this interaction was not included in the data set.

The social network graph of all study participants across all Notre Dame cohorts, based on beacon interactions, indicates that there is substantial structure in the data. Each of the cohorts 1-4 registered as a distinct cluster in the graph, and cohort 5 was scattered around the graph, which is consistent with our understanding that it was recruited from the organizations already represented in cohorts 1-4. Connections between some of the cohorts suggest that some of them may be co-located or organizationally connected.

To identify groups as they persist over time, we analyzed the graph of interactions using a Louvain clustering algorithm. To capture the time element, we created a series of interaction graphs, each reflecting the interactions between participants over one specific week of the study period. Simply running one clustering algorithm on each disjoint weekly graph, however, provides no consistent way to link clusters from one week to clusters from another. To address this issue, we followed a multilayer graph analysis approach, in which we added self-edges to connect the same participant across different time windows. We then used a clustering algorithm on this large, multi-week graph to identify clusters that encompass multiple time windows. Because of the self-edges, the algorithm has a preference to keep a participant in the same cluster over time, which provides continuity in clusters from week to week.

Having identified groups that persist over time, we were then able to characterize some of their dynamic features. In examining the start and end dates of groups, we noted a possible transition that takes place around week 20 of the study in both cohort 1 and cohort 2, where several groups come to an end while several new groups take their place. This suggests some type of organizational change or event may have occurred in this time frame. In addition, we analyzed gender composition and stress levels of groups over time, although no generalizable patterns were apparent in these analyses.

To characterize influence relationships among participants, we modeled influence at both the population level (i.e., a person's overall influence on everyone they interact with) and at the level of interactions between individuals (i.e, a person's distinct influence on each person they interact with) using three different influence modeling approaches: A generalized linear regression mode, a lasso regression model, and a Markov model. We define influence as the ability of a person to influence the behavior or psychological state of those they directly interact with. At a population level, we used the models to identify a group of particularly influential individuals across all cohorts. We were then able to characterize this group of influencers in relation to the general population on several metrics. Our results indicate that influencers are often less central in the interaction graph than other study participants, which is somewhat counterintuitive. Influencers are also, in general, less likely to use tobacco and more likely to drink alcohol than other participants. There are also more subtle patterns in personality traits of influencers that are more difficult to interpret. Based on the lasso regression model, we were able to compare the population-level influence of each person across multiple ground truth measures, including positive affect, negative affect, alcohol use, work behaviors, and sleep behaviors. This analysis demonstrates that individuals who are influential on one behavior tend to be influential on multiple behaviors, which suggests some consistency in their ability to influence others. We were also able to use the individual interaction-level model to develop graphs showing influence patterns among individuals throughout each cohort.

In addition to our analysis work, we describe a framework for developing a MOSAIC data navigation capability. To develop this framework, we created a MOSAIC knowledge model based on general concepts in signal analysis, as well as concepts related to domain-specific sensors and signals for the MOSAIC project. To explore ideas for a navigation interface, we translated the model into JSON format to use as a stub for a persistent data tier in an actual application. We also implemented a basic version of a hierarchical data browser interface in JavaScript.

In conclusion, we discuss several implications of our work. First, we discuss limitations of our analysis, in particular uncertainties that may be introduced by the asymmetry of interaction durations between beacons and the possible false negative issues in the Notre Dame processed interaction sessions data. We then make several recommendations for future data collection and management to address sampling issues, the need for additional organizational ground truth, possible sensor reliability issues, and ways of better enabling data sharing. Finally, we present a number of proposals for testing hypotheses based on our analyses, focusing on organizational characteristics, group formation and dynamics, and influence.

CONTENTS

1 INTRODUCTION

This report is the result of a 4-month project to analyze data provided to Los Alamos National Laboratory (LANL) by performer teams for the Multimodal Objective Sensing to Assess Individuals with Context (MOSAIC) project. The MOSAIC performer teams monitored hundreds of volunteers in their daily activities, using wearable sensors to capture data about interactions, physical activity, physiological state, and other measures of interest. The project also collected "ground truth" data from participants in an initial survey and through short daily surveys through the study, providing insight into their day-to-day psychological status and behavior. The original goal of the project, as we understand it, was to understand whether wearable sensor data could be used to predict individual behavior and mental state over time. The LANL MOSAIC work presented here has a slightly different goal, which is to use both sensor and ground truth data to capture and analyze social dynamics within organizations. To do this, we focus on two key objectives: identifying and characterizing groups within organizations as they evolve over time, and identifying and modeling patterns of social influence among participants. This has enabled us to identify interesting patterns in the data, and to develop a set of hypotheses that could be tested against additional ground truth data. We also report on some of the challenges we faced working with the data, and provide recommendations for future data collection efforts aimed at organizational dynamics.

In order to apply the broadest possible base of expertise to the project, we managed it in two stages: an intensive data sprint to jump start analysis efforts, followed by a longer-term effort to extend the most promising analysis strategies. The sprint was conducted over three and one half days, May 13-16, 2019, and was facilitated by LANL in cooperation with Air Force Cyberworx, an organization with significant design sprint experience. The sprint was attended by 15 LANL participants and three outside participants (including two from the Lockheed Martin team), representing research disciplines ranging from social and biological sciences to mathematics and computer science. Over the course of the sprint, participants engaged in an array of planned activities designed to generate and refine hypotheses to be tested in the data, followed by a short an intensive period of analysis. On the final day of the sprint, the IARPA project manager attended and was briefed on the results. Although arranging data access and organizing the sprint was a more complex task than initially anticipated, the sprint was generally judged to have been a success and a possible model for future analysis efforts at LANL. Of particular note was the use of Amazon Web Services (AWS) computing resources for the sprint and subsequent analysis efforts, which enabled rapid deployment of customized computing resources tailored for the project.

For the extended analysis phase of the project, five of the sprint participants agreed to work together to extend their analyses from the sprint. In addition, two other participants continued work to provide data navigation concepts and tools. Based on successful analysis strategies from the sprint, this group decided to focus on influence modeling and group identification and dynamics. In order to make these analyses possible, the team first spent a large portion of their time characterizing the data, identifying data that were usable and relevant to the analysis tasks, and processing data to prepare it for analysis. Through these efforts, we identified the data set provided to us by the Notre Dame team as the most readily accessible for analysis [9]. This was largely due to the fact that the Notre Dame team provided their data in the form of a well-structured, readily navigated database, and had already performed some data pre-processing. In our judgment, the datasets provided by the Lockheed Martin and University of Southern California teams, though of immense interest, were too voluminous and in too raw of a form to be amenable to in-depth analysis within the short time span of the project. In addition, within the Notre Dame data, we determined that the beacon data was most relevant to our particular analysis goals, as it provided a direct proxy for face-to-face social interaction. We also made significant use of ground truth data in our influence modeling, which is aimed at showing how mood and behavior are mediated by social interactions.

One major challenge for this work is the lack access to organizational ground truth. While we do have access to ground truth measurements of individual psychological and behavioral patterns, we were given very little information about the organizations in question. Aside from a few pieces of information provided by IARPA and the MOSAIC performers, we lack information on the size, structure, or type of work engaged in by the organizations, and have very little information about the roles or locations of individual participants within these organizations. In addition, with the exception of one cohort of participants, we do not know what fraction of employees were sampled within each organization, or whether the samples were concentrated within certain work groups or spread more widely across the organizations. This makes interpretation of results difficult, because in a typical organizational analysis project, substantive findings are driven by analysis of network structures in relation to known organizational features. Without this ground truth, we have no clear basis for validating or interpreting our results in a real-world context. While we make extensive use of existing ground truth data on participant demographics and psychological/behavioral status, which addresses this issue to some extent, we would be able to refine and validate our analysis to a much greater degree with access to additional organizational ground truth. The list of proposals for testing our results against ground truth data in section 7 provides a starting point for how we could approach future validation and refinement efforts.

The remainder of this report is structured as follows: Section 2 provides an overview of the Notre Dame beacon data, with a particular focus on unexpected asymmetries in interaction time between participants as captured by their beacons. Sections 3 and 4 describe the methods we use to identify groups within the data that are continuous in time, and examine group dynamics and characteristics over time. Section 5 describes the results of three different influence modeling approaches applied at both population and interaction levels, which enables us to identify and characterize the population of influencers within each organization. Section 6 describes a knowledge model for the MOSAIC data and initial efforts to develop a data navigation tool. Finally, Section 7 presents our conclusions, including recommendations for future data collection and a set of hypotheses that might be testable against organizational ground truth.

2 CHARACTERIZING BEACON DATA

The complete dataset collected by the the Notre Dame team included 757 participants, recruited in five cohorts. Our early phase of methodological development concentrated on the 26 study participants recruited into Cohort 1. Documentation indicates that these participants work together at a single location of a small computer software company, and that all or most of the employees at this location were part of the study. These features made this cohort a useful place to start characterizing the interactions captured in the beacon data, because they indicate that all of these participants are in fact likely to be in regular physical proximity to each other. For other cohorts, we understand that various recruitment strategies were employed that could result in uneven sampling that could complicate interpretation of the data. Our full analysis effort did cover other cohorts, as described in later sections of this report. Table 1 summarizes some basic demographic characteristics of the 21 unblinded participants in Cohort 1, i.e. those for whom we have ground truth data.

ID	Hashed ID	Age	Education	Occupational Category	Sex	Reports
01	9780495721948428633	26-30	College degree	Computer & Math	M	—
02	9883624450358368517	20-25	College degree	Office & Admin Support	F	—
03	10104865842667159622	20-25	College degree	Computer & Math	F	—
04	10267235808015800433	36-40	College degree	Computer & Math	M	5-15
06	10439117077451641565	30-35	College degree	Computer & Math	M	—
08	10806377045013041963	26-30	College degree	Computer & Math	M	—
10	11049835262840569530	46-50	Some college	Computer & Math	F	—
12	11325750135091548264	30-35	Master's degree	Computer & Math	F	—
14	11439412180368165013	30-35	Master's degree	Management	M	5-15
15	11750352889969585637	40-45	College degree	Computer & Math	M	—
16	11874486772857607558	20-25	Some college	Computer & Math	M	—
17	12115852993242340806	56-60	College degree	Computer & Math	M	—
18	12204601337084291188	36-40	Doctoral degree	Business & Financial	F	—
19	12206210069307053438	26-30	College degree	Computer & Math	M	—
20	12502660685247488411	26-30	College degree	Computer & Math	M	—
21	12718614204893805965	26-30	College degree	Computer & Math	M	16-30
22	12893446853201738411	36-40	College degree	Computer & Math	M	—
23	12994028865111584912	26-30	College degree	Computer & Math	M	< 5
24	13094498676145090523	20-25	College degree	Computer & Math	M	—
25	13146988918954476898	26-30	College degree	Other (unspecified)	F	—
26	13339207273792338820	30-35	Some college	Computer & Math	M	—

Table 1. Demographic attributes of unblinded participants in Cohort 1. Participants with supervisory roles are indicated by an entry in the "Reports" column, indicating the number of people reporting to them.

2.1 Asymmetric interaction durations

During development and evaluation of the influence modeling methods detailed below, we noted a counter-intuitive property in the interaction data. Contact time t should be transitive between any two people A and B: $t(A, B) = t(B, A)$ because it implies mutual proximity. We noted that this property did not generally hold true in the Cohort 1 beacon data (Figure 1). Analysis of both the raw beacon data and interaction sessions the Notre Dame team computed from those data provide insight about the intransitivity of contact times, and suggest alternative ways to acquire and analyze beacon data to quantify interpersonal interactions.

To investigate the asymmetry shown clearly in Figure 2, we use data from beacon interactions captured after February 25, 2018, in order to screen out potential irregularities in data collection during initial beacon distribution. Starting later in the sampling period (the last enrollment date was April 6, 2018) does not

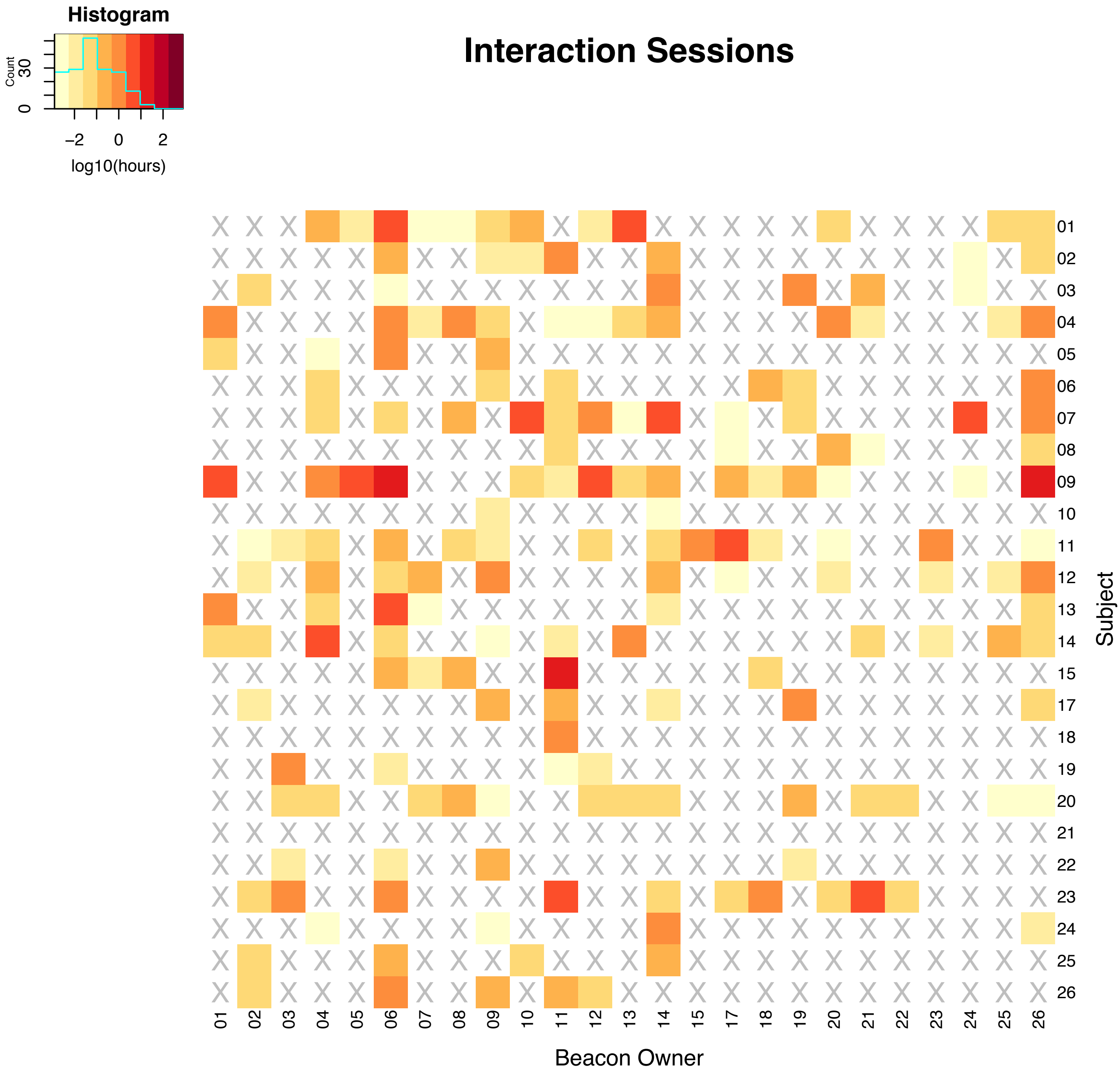

Figure 1. Heatmap of cumulative interaction times between participant pairs. For each pair, the owner's beacon, shown column-wise, emits a signal that is received and logged by the mobile phone application of another study subject, shown row-wise. An X indicates missing data, presumed to indicate an interaction time of 0. If interaction times were transitive, this chart would be symmetrical around the diagonal from top left to bottom right. Data shown are from processed interaction sessions, rather than raw beacon interactions; the latter show similar patterns, shifted toward higher values, and include self-detection records. Data were filtered for face-to-face interactions of 5 seconds or more, as described in the text.

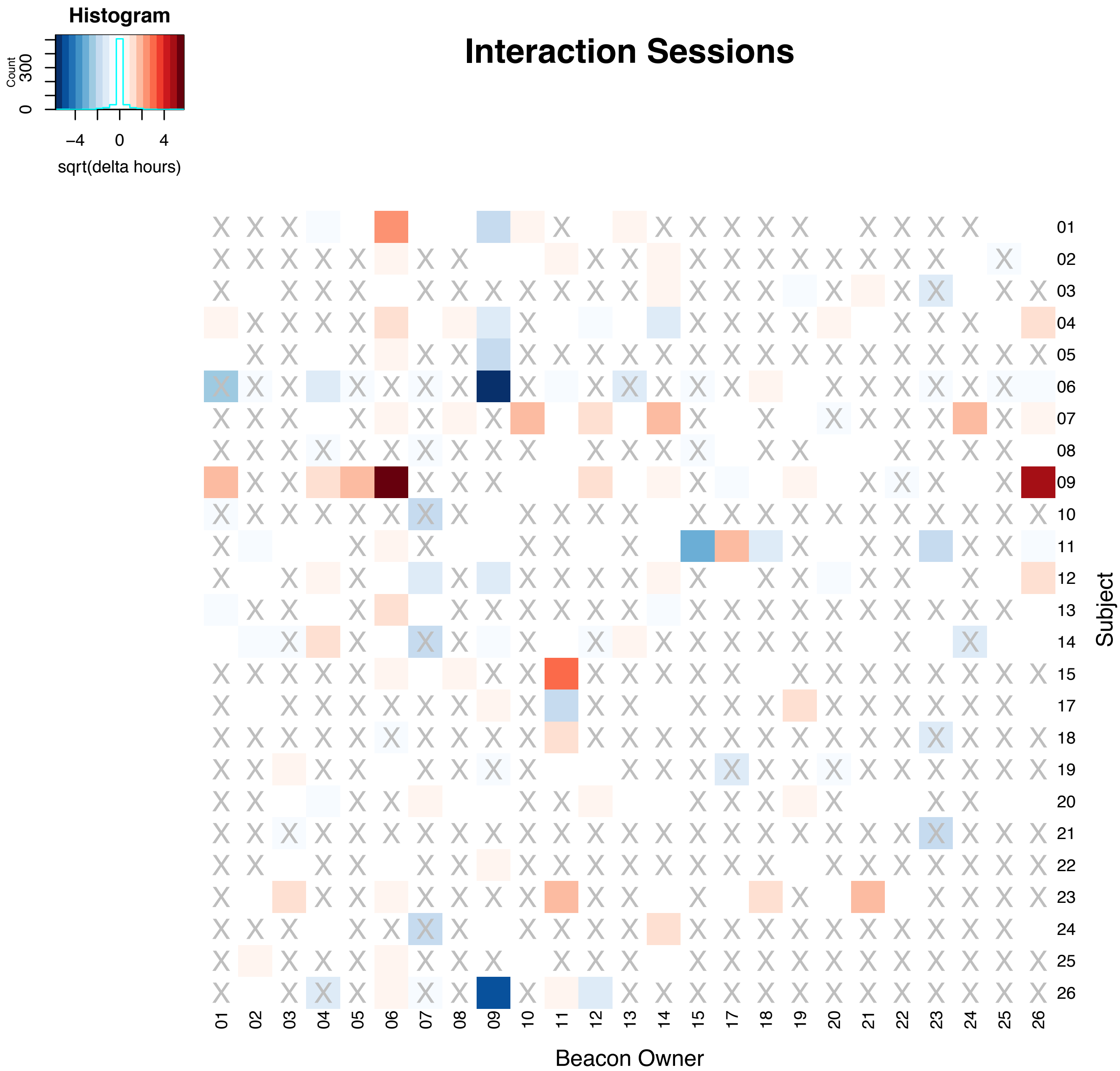

Figure 2. Interaction duration asymmetry, from the same data used in Figure 1. Heatmap shows the difference in cumulative interaction times between participant pairs, $t(i,j) - t(j,i)$. An X indicates missing data, again presumed to indicate 0 interaction time. Differences were square-root transformed (preserving sign), to better depict both large and small deviations.

change the results significantly. We further filter the data to require a minimum of 5s interaction duration and threshold (or peak) beacon signal intensity of -60 dB, to remove transient and distant signals. We also exclude the one participant who dropped out of the study for reasons unknown.

Reviewing the full beacon data for Cohort 1 reveals some possible non-compliance. Two participants (10 and 13) seem never to have detected any of their beacons and five (10, 13, 20, 22, 25) never detected their own mobile beacons (not shown in these figures). The cause is unclear, because all of the above participants were detected by other participants' mobile phone apps and also detected others' beacons with their phones. Participant 21 (a high-level supervisor) logged no beacon interactions with their phone receiver application.

2.2 Raw beacon interactions versus processed interaction sessions

Here, we compare the raw beacon interaction times with a table of pre-processed "interaction sessions" calculated by the Notre Dame team, both of which are included in the postgres database provided by Notre Dame. Comparison of raw mobile beacon interaction times with the processed interaction sessions requires that both versions of the data are screened similarly to remove interactions shorter than 5s and weak signals below -60 dB, and sampled for the same calendar period. We compute total interaction time in hours for each possible pair of individuals from two different data sources. The processed beacon sessions include self-detection of mobile beacons (wallet and keychain), which we excluded. The raw beacon interaction data includes duplicated detection events and overlapping interaction sessions. We consolidate overlapping beacon sessions for accurate contact times, so as not to overcount time from overlapping and duplicated interaction sessions. Despite having consolidated the raw beacon session data into non-redundant, non-overlapping intervals, the raw data show far greater interaction times, 2620.3 hours in total, than the processed interaction session data (Figure 3), with 147.2 total interaction hours. The asymmetry in total interaction times per participant is evident even in this aggregate view, in both the raw and processed data.

For Cohort 1 participants, total interaction time logged by mobile beacons is uncorrelated with the total interaction time detected by the mobile phone beacon-detector application (Figure 4). This is surprising, because interaction time should be an intransitive property of an interaction, and ideally should not differ for the beacon-detector pairs used to measure it. In practice, small inconsistencies most likely contribute to large differences, which could include situations where a mobile beacon and a phone receiver are not both carried together routinely, the mobile phone receiver application is not fully enabled to capture and record beacon signals, or beacon batteries are depleted. The raw beacon data might be used to evaluate compliance by identifying intervals in which both beacon and mobile phone receiver are together in proximity, which appear as self-detection events in the interaction logs, then conditioning analysis of interaction sessions given only completely instrumented intervals. This would yield less data, but with fewer false negatives.

It is likely that false negatives are prevalent in these data. If each participant spent even an hour a week in face-to-face interaction with others during the 38-week study period, we would see 38 hours of interaction per person. The interaction sessions data show much less interaction than this, capturing an average of 5.9 hours of interaction time per participant (median 2.2 hours, range 0 to 68.8 hours). Even the individual who spent the most time with coworkers (09) recorded 68.8 total interaction hours, which averages out to 1.8 hours per week, roughly 21 minutes per work day, spent interacting with others (of that, 33.4 hours – 48.5% – was spent with one other participant, 06). These totals seem low based on our organizational experience.

The raw mobile beacon session data show much more interaction time than the Notre-Dame processed interaction sessions data, recording a mean of 104.8 interaction-hours (median 55.3 hours, range 6.2 to 583.7 hours) per participant. The most interactive individual (09) spent an average of 15.4 hours per

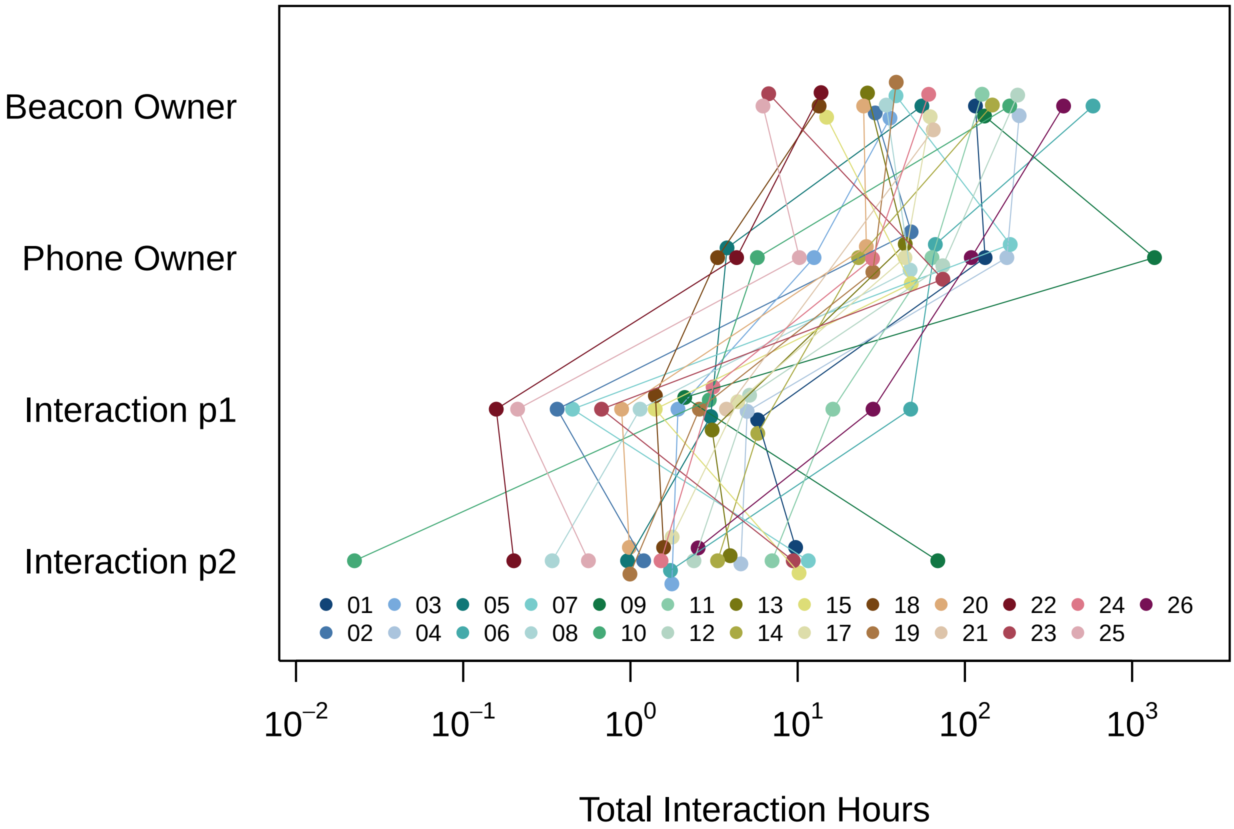

Figure 3. Total interaction hours recorded per participant, according to different data sources. Symbol color uniquely identifies each participant; lines are included to make it easier to track differences in participants' measured interaction durations across data sources. For raw data on beacon sessions, cumulative time that each beacon was detected by a phone ("Beacon Owner") is shown separately from the total time that each phone detected a beacon ("Phone Owner"). Similarly, for the processed interaction sessions, the two partners in each interaction (p1, p2) are shown separately. Values for p1 interaction totals are the column totals from data in Figure 1 and p2 are the row totals. The differences in cumulative interaction times are related to the asymmetric interaction durations of concern.

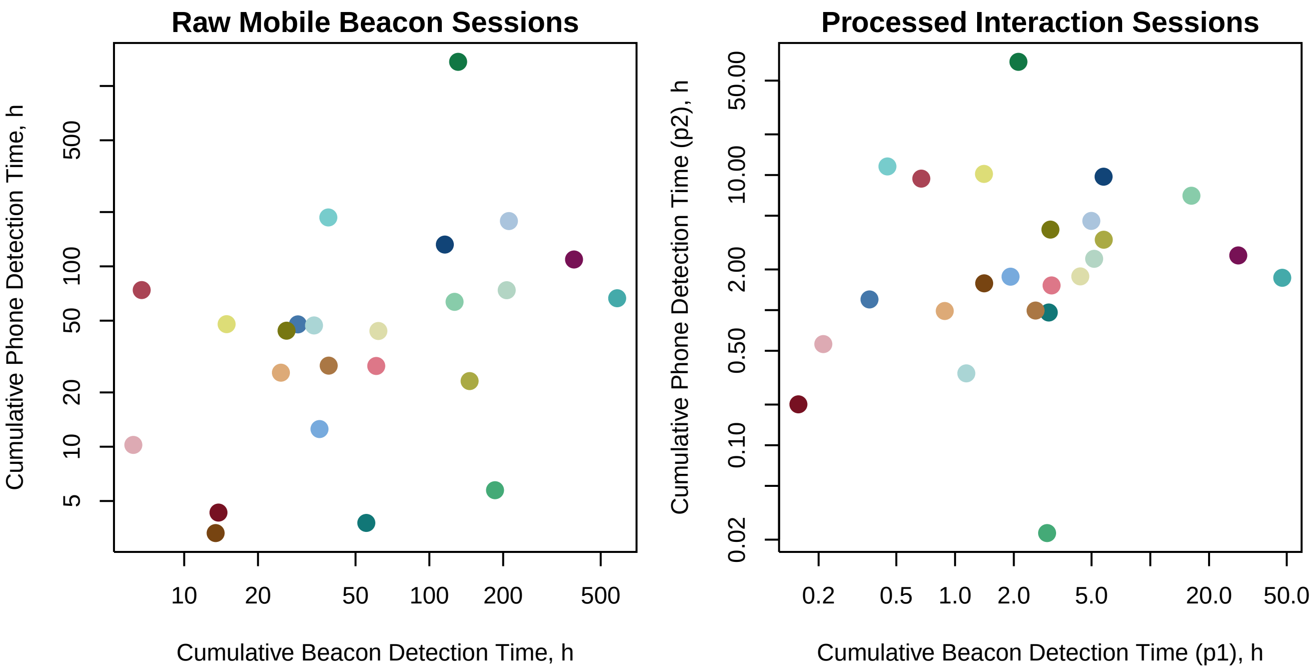

Figure 4. No evidence for significant correlation of total detection times (hours h) between a participant's mobile proximity beacons and phone detector. Each point indicates one participant, with colors and data as shown in Figure 3. This view compares the total duration of interaction sessions for each participant's mobile beacons, as detected by other participants (x-axis), with the total time each participant's mobile phone application detected beacons of other participants (y-axis). The left panel summarizes raw mobile beacon detection data consolidated into non-redundant, non-overlapping intervals (Kendall's $\tau = 0.233, p = 0.108$). The right panel shows the processed interaction sessions (Kendall's $\tau = 0.225, p = 0.131$). These data indicate that beacons and phones are mutually independent, while the practical approach of taking the maximum interaction duration between each pair of participants induces a perfect correlation.

week over the 38-week study interacting with others. This may be a more realistic amount of interaction. However, the raw mobile beacon data are likely also to include false negatives because of the asymmetry in interaction times logged by transmitters and receivers. These discrepancies suggest that the Notre Dame team excluded some interactions when processing the interaction sessions data, based on criteria not known to us.

For future data collection efforts, it may be helpful to test beacon detections for asymmetries before beginning data collection, or to document differences in beacon sensitivity. Adopting a beacon platform that combines transmission and reception in a single device may also improve sampling consistency. In addition, email or phone communications data could be used to augment beacon data. The methods described below would work just as well on such enhanced data as they do on the data summarized here.

In order to avoid inconsistencies, for the analyses described in the remainder of this report, we force interaction times to be symmetrical by using the maximum interaction time recorded by either member of each pair of participants, as obtained from the processed interaction session data.

3 GROUP DETECTION

3.1 Building a social network graph

To identify dynamic groups of individuals within the data provided by Notre Dame, we first need to define a social network in the context of the available data. We use the pre-processed "interaction sessions" table in Notre Dame's postgres database as our basis for defining an interaction, which in turn defines an edge connecting study participants within a social network. The "interaction sessions" data contains entries detailing which two people were in proximity to each other, as measured by their mobile (wallet or keychain) beacons, the start and stop time of the interaction, the strength of the interaction signal, and the duration of the interaction in seconds.

In building the social network graph, we make the following choices:

- Analyze participant cohort 1 through 4 separately, based on guidance from Notre Dame performers that each cohort represents a different "organization." We do not analyze cohort 5, as these participants were recruited by participants across all other cohorts, and have no inherent connection to each other.
- Limit interactions we consider to those with duration greater than 0 seconds and with an RSSI Threshold of -60, which is roughly defined as a face-to-face interaction.
- Only include participants who are not blinded in the available data, so that we have the ability to consider their demographics and psychological metrics in combination with social network structure.
- Assume that all interactions are symmetric. Due to idiosyncracies of participant compliance, reliability of beacon sighting measurement, or other artifacts of the data, some interactions are reported in the raw data as non-symmetric, as described above. That is, in some cases, participant A reports having interacted a different amount of time with participant B than participant B reports having interacted with participant A. However, common knowledge tells us that the real interaction time must be at least the maximum of the two reported durations. Therefore, in any case where a reported interaction is asymmetric, we use the maximum of the two recorded durations.

The full network over the entire study period is shown in figure 5. Each node in the network represents one person (node color indicates which cohort they belong to) and edges indicate some face-to-face interaction between two persons (nodes). The underlying data set that informs the graph includes the total interaction time for each pair of persons, but that is not represented in the image. Just from examining the plotted network we can see there is substantial structure to the data. Without knowing that there were 5 cohorts we can clearly see 4 distinct communities corresponding to cohorts 1-4, but we might wrongly assume that cohort 5 is not a separate enrollment because they co-cluster with existing groups. Likewise, with the cohort ID information we can see substantial connections between cohorts 2 and 3 and even stronger overlap between cohorts 2 and 4, suggesting that cohorts 2 and 4 share one or more physical locations.

While figure 5 implies significant and interesting network structures exist in the data, many of the links between nodes are fairly weak. Figure 6 shows the same network, but with edges representing less than 1 hour of cumulative interaction time removed. Even in this very large network, significant sustained social interactions are rare, with many occurring within the small cohort 1. This sparsity is likely caused by differential sampling in the different cohorts. Based on our understanding of the data collection process, participants in cohort 1 were all co-located employees of a small company, meaning that beacons could have captured most of their workplace interactions. In the larger cohorts, the data may represent a smaller fraction of employees within each organization, meaning that they also capture a smaller fraction of total

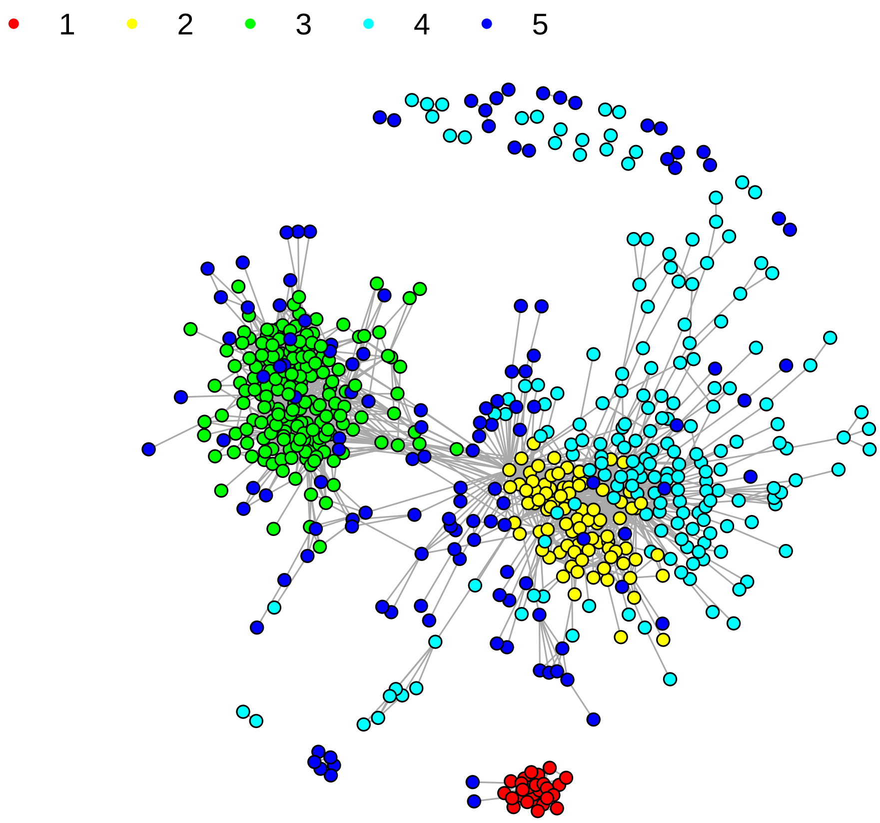

Figure 5. Network of all face-to-face interactions in all five Notre Dame cohorts. Nodes represent individual people, colors indicate the cohort ID and edges indicate at least one face-to-face contact over the course of the whole study period.

interactions for each participant, since participants are likely to interact with coworkers who are not enrolled in the study.

3.2 Clustering

After constructing a social network graph for each of cohorts 1 through 4, we now attempt to detect social groups (i.e., *clusters*) that exist within the social network. In order to have a stronger basis for analyzing group dynamics, however, we need to be able to define and detect these clusters across time, not just as they exist in a static network. This requires splitting up the graph into a series of time slices, each of which is a graph representing all the interactions occurring in a time window of pre-defined size. Due to the sparsity of available data, we use a time window of one week. Then, we create one social network graph per cohort per week (using the choices described in the previous section) and weight the edges by the total interaction durations over that week.

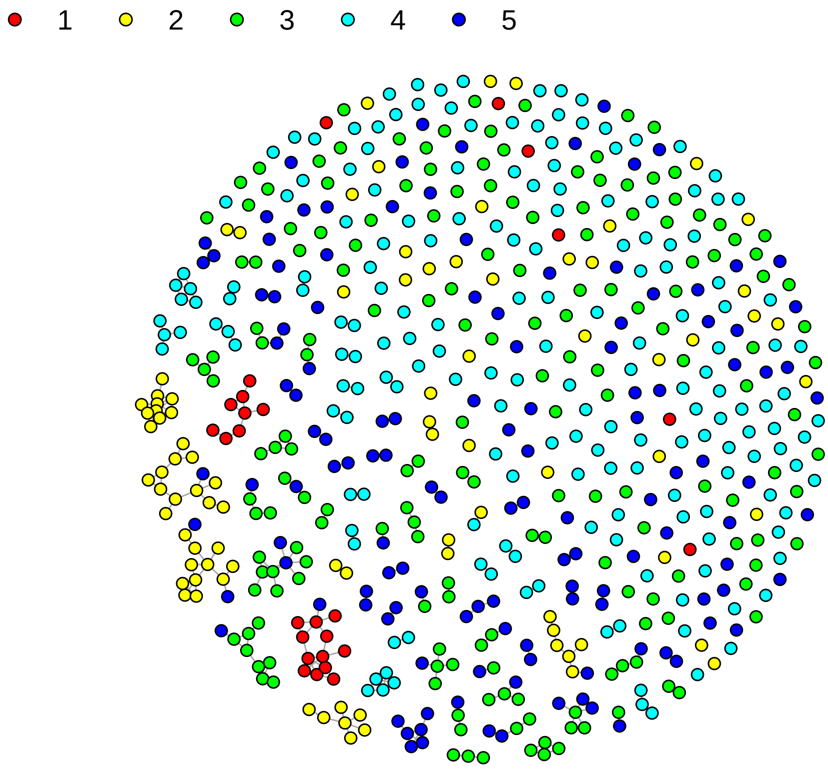

Figure 6. Network of all face-to-face interactions greater than one hour in duration in all five Notre Dame cohorts. Nodes represent individual people, colors indicate the cohort ID and edges indicate at least one or more hour of face-to-face interaction over the course of the whole study period. Note that this is one hour of aggregate interaction time over the whole study period and not necessarily one continuous hour of interaction.

To find clusters that persist across time, we cannot simply run one clustering algorithm on each disjoint weekly social graph, because it would be impossible to identify the same cluster existing across time windows. For this reason, we follow the technique proposed in Mucha *et al.*, where we build one large social graph that spans across the entire study, adding self-edges to connect the same participant across different time windows [10]. The final large graph contains one node for each (participant, time window) pair, and in addition to edges connecting different participants, it also includes edges connecting a participant in time window *t* to him/herself in the next time window *t+1*.

Figures 7 – 9 illustrate how these elements of the graph come together. Figure 7 shows the full, "unrolled" interaction network for cohort 1, showing numerous disconnected networks, each corresponding to the interactions in cohort 1 in a given week. Figure 8 shows the same network, but only includes the edges linking one person in one week to themselves in another week (self-edges). Figure 9 shows the combination of the unrolled network showing both the weekly interaction edges and the self edges.

Even for a very small cohort, the full, unrolled network at the week resolution level with both interaction and self-edges is complex and unwieldy. However, with this network (Figure 9) we can now re-cast the intractable problem of identifying dynamic groups from beacon interaction data as the tractable problem of detecting community structure in this static, albeit complex, graph. To do this, we consider the weight on the self-edges as a parameter that we can tune. The clustering algorithm is then able to cluster across all time while still looking at social structure on a weekly basis, and has a preference (based on the selected self-edge weight) to keep a participant in the same cluster across weeks, unless the local structure around the participant shifts significantly.

After building one of these larger graphs containing self-edges for each cohort, we can then apply any clustering/community detection algorithm to detect groups. For this study, we use Louvain clustering, which relies on a single resolution parameter, and already has an easily useable implementation in Python. Note that Louvain clustering is a hard clustering algorithm — it assigns each participant to exactly one cluster per time window. Future work could investigate other clustering algorithms and relax the hard clustering restriction.

Having designed our representation of the social network structure over time, and chosen a clustering algorithm to detect social groups, we then run a sensitivity analysis to chose values for the self-edge weight and clustering resolution parameters that seem reasonable. Because we do not have access to organizational ground truth against which to evaluate our detected clusters/social groups, we choose "reasonable" parameters via subjective human evaluation of four metrics: (1) the number of singleton clusters (clusters which only contain a single participant) discovered, (2) number of size > 1 clusters (clusters which contain more than a single participant) discovered, (3) the average number of weeks over which a cluster persists, and (4) the average number of clusters a participant belongs to over all time. Generally, we want to minimize the number of singleton clusters, discover a manageable number of size > 1 clusters, find clusters that persist for more than one week but less than the whole study time period (because we are interested in group dynamics), and assign each individual to 2-3 clusters over the course of the entire year. We these guidelines in mind, we choose appropriate parameter settings for each of the four cohorts. The characteristics of the discovered clusters are discussed in the next section. For details on our sensitivity analysis, see Appendix A.

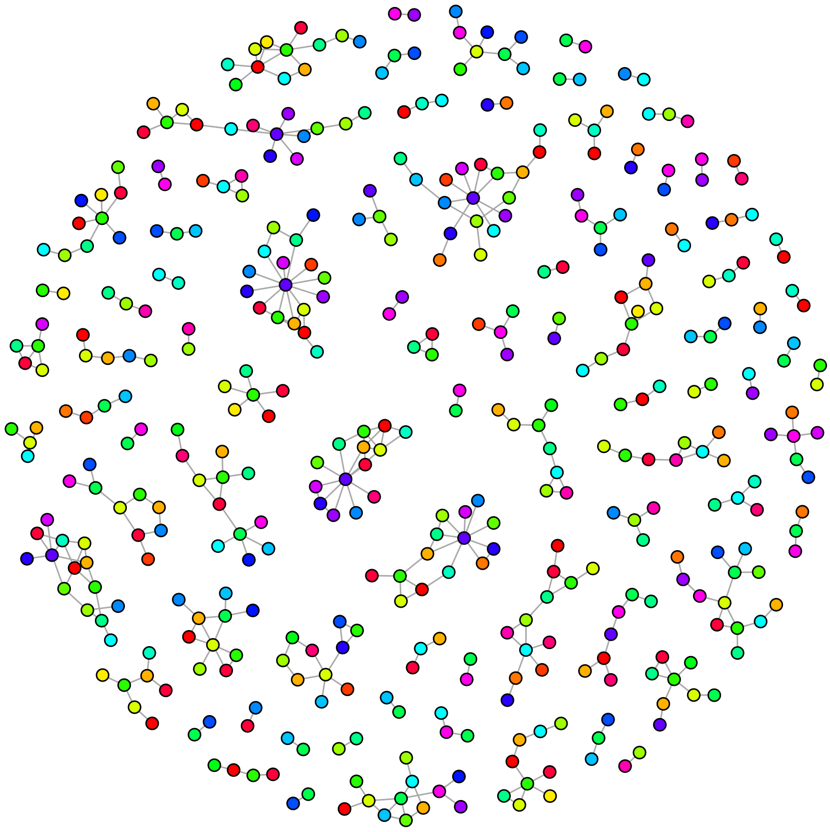

Figure 7. Weekly interaction graph of cohort 1. Each individual is represented by a single color, with multiple nodes of the same color indicating that the individual was part of the graph for multiple weeks. Connected subgraphs represent interactions taking place within a give week of the study period.

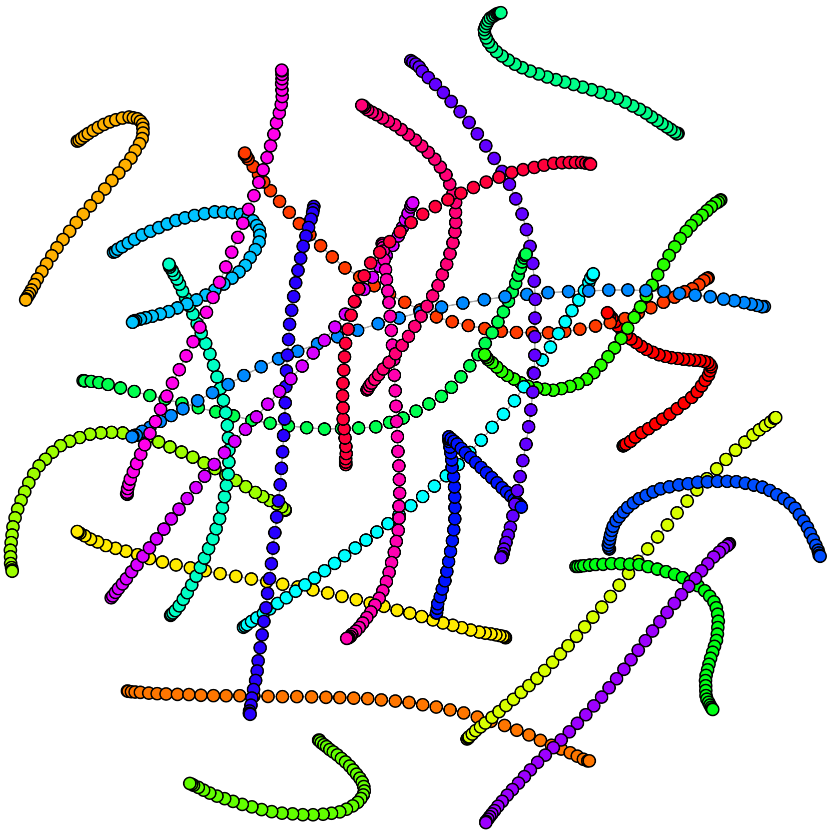

Figure 8. Weekly self-edge graph of cohort 1. Each set of nodes of the same color again represents a single individual, but here only the self edges that link the same individual from one week to the next are shown

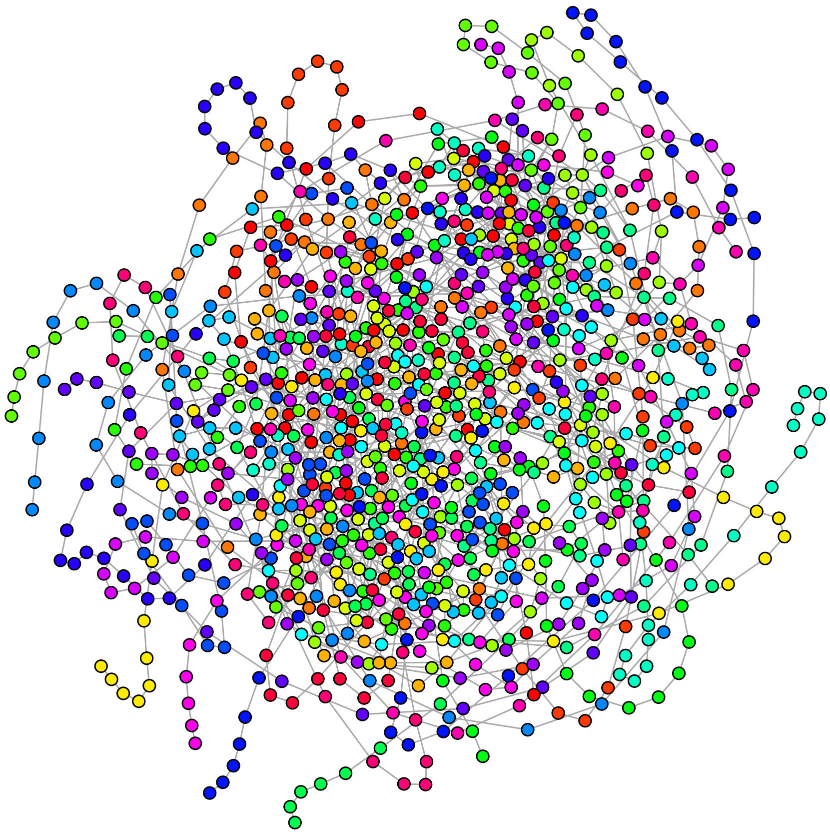

Figure 9. Weekly interaction graph of cohort 1 with self edges included. This graph includes all of the edges present in both of the previous graphs. Each set of nodes of the same color again represents a single individual.

4 GROUP CHARACTERISTICS AND DYNAMICS

Based on the guidelines above, we implement Louvain clustering on each cohort's social network graph, using the following parameter settings: for cohort 1, cluster resolution is 3 and self-weight is 800, for cohort 2, cluster resolution is 3 and self-weight is 400, for cohort 3, cluster resolution 2 is and self-weight is 500, and for cohort 4, cluster resolution is 2 and self-weight is 3900 (the average interaction duration is much larger in cohort 4). Also note that we do not have a way to account for missing data due to sampling rate of the population, participant non-compliance, etc. We consider clusters that contain more than one participant and that also change size over time as "interesting," and specifically investigate those further.

4.1 Longevity

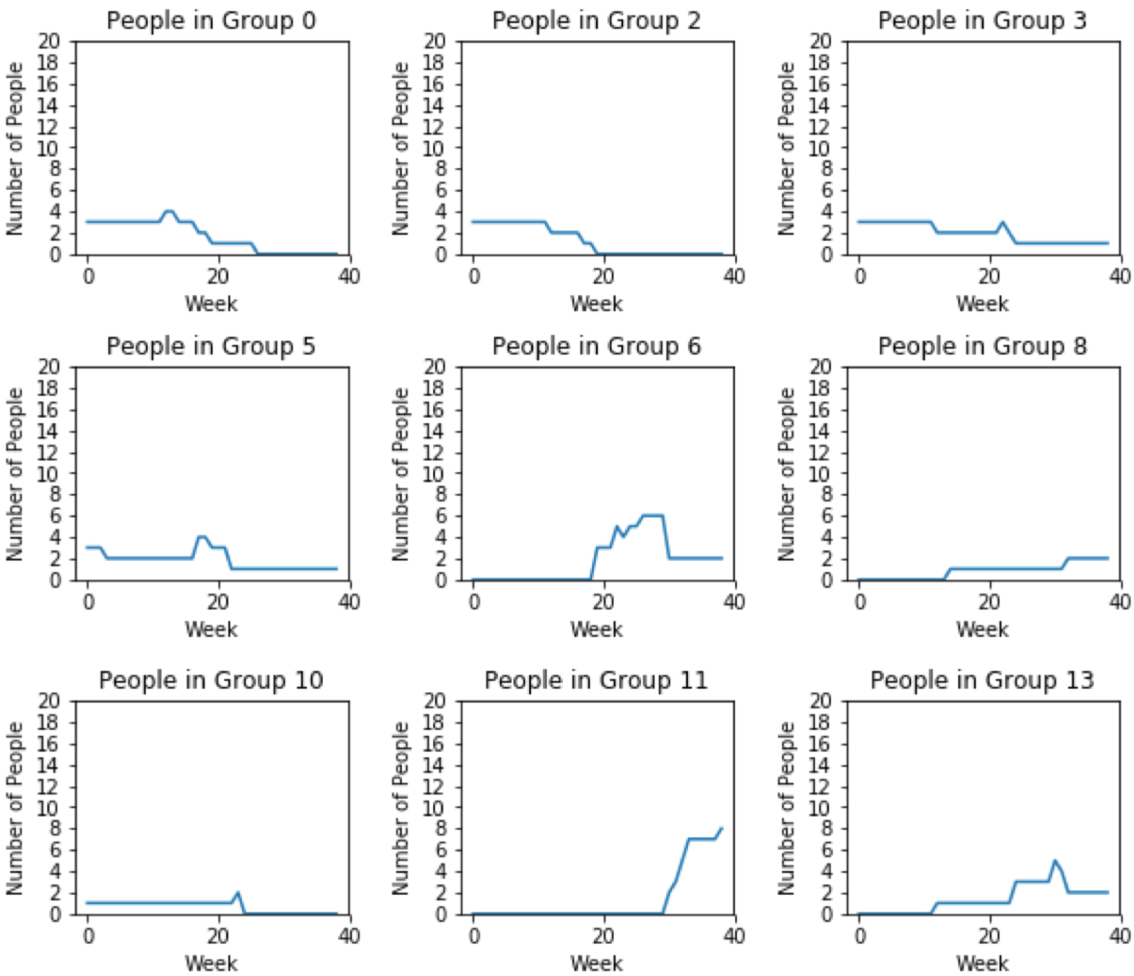

Figure 10. Sizes of interesting groups in Cohort 1 over time. Note the change around week 20.

In cohort 1, we find 9 social groups with more than one member and which change over time. Figure 10 shows the size of these groups over time. There appears to be some kind of change around week 20, about halfway through the study. In particular, groups 0, 2, and 10 seem to decline in size and disappear, while groups 11 and 13 form and persist. Also, group 6 appears to form around week 2 and decline in size around week 30.

In cohort 2, we find 15 social groups with more than one member and which change over time. Figure 11 shows the size of these groups over time. Some groups have a notable change, again around week 20. Particularly, groups 0, 1, and 6 seem to form around week 20 and persist.

In cohort 3, we find 26 social groups with more than one member and which change over time. Figure 12 shows the size of the most interesting groups over time. Group 1 stays consistently large, while group 26 steadily loses members. Group 47 forms and dissolves. Groups 5, 6, 9, and 12 consistently increase in size, and group 15 fluctuates.

In cohort 4, we find 5 social groups with more than one member and which change over time. Figure 14 shows the size of these groups over time. We do not find any particularly interesting dynamics to investigate, because no group is ever larger than 2 people.

4.2 Gender distribution over time

For each group in cohorts 1-3 we look at the gender distribution over time. Because our clustering on cohort 4 did not discover any particularly interesting groups, we no longer consider cohort 4.

Figure 15 shows the gender distribution of groups on cohort 1. Groups 0, 3, 8, 11, and 13 see significant changes. At the same time, groups 0 and 3 decrease in size while groups 8 and 11 get larger.

Figure 16 shows the gender distribution of groups in cohort 2. Group 0 shows interesting behavior. The others that show changes in male percentage only have 2 people. During the same time, groups 0 increases in size, implying that it was initially more female and then became more balanced.

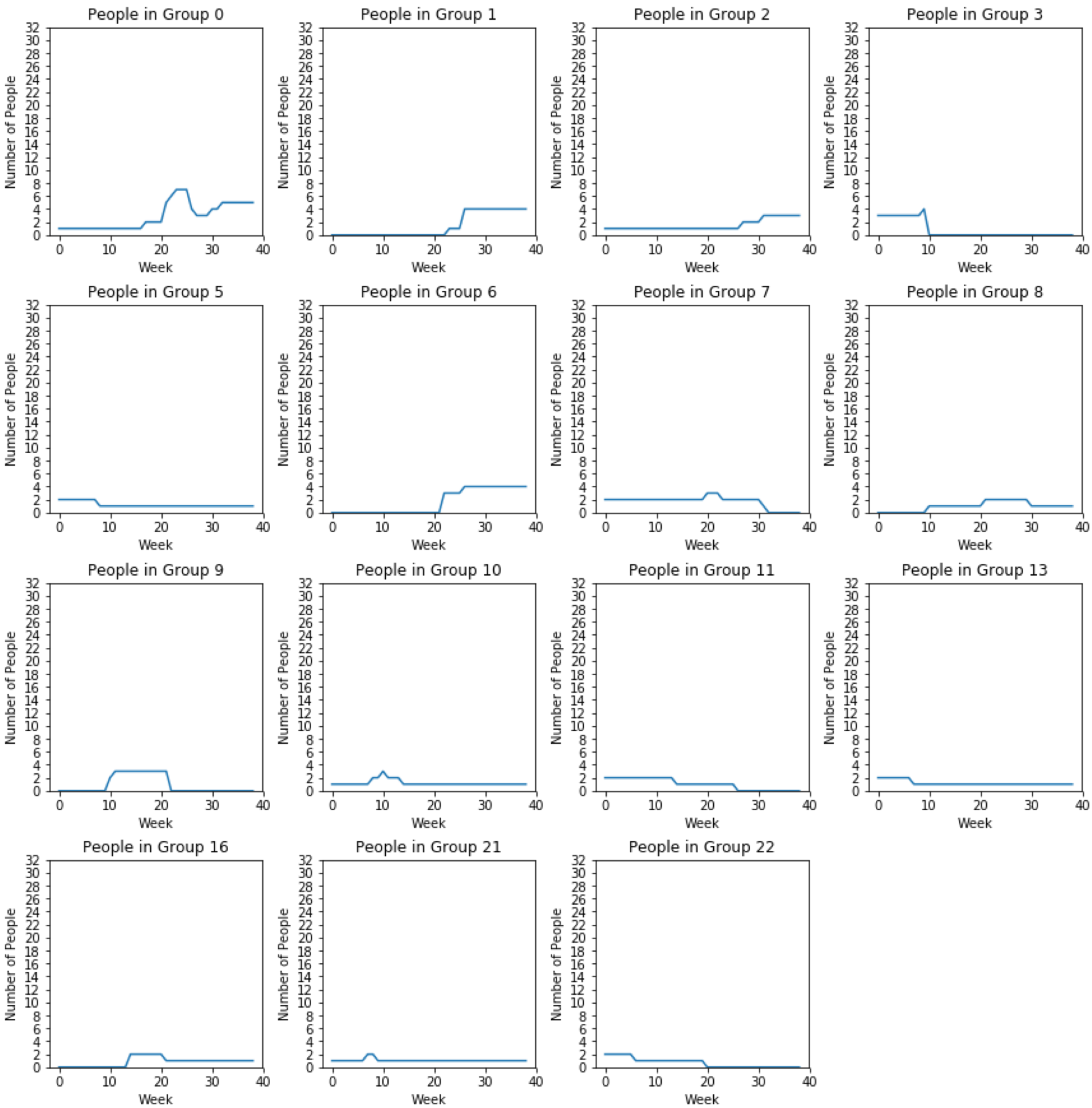

Figure 11. Sizes of interesting groups in Cohort 2 over time. Note some changes in groups 0, 1, and 6 around week 20.

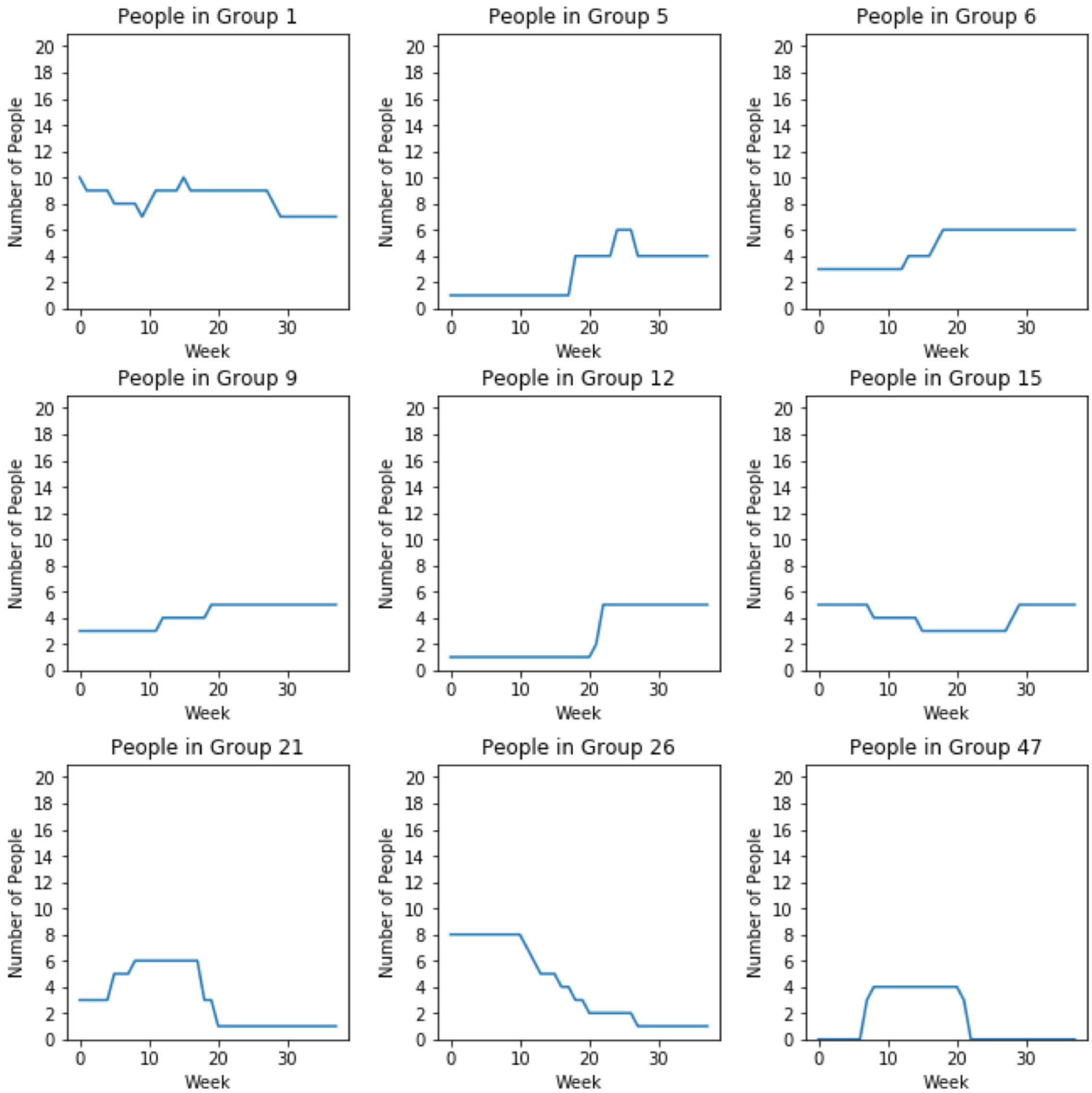

Figure 12. Sizes of interesting groups in Cohort 3 over time. Note some groups increase in size while others decrease, some rapidly and some more slowly. Group 15 loses some members, and then gains them back.

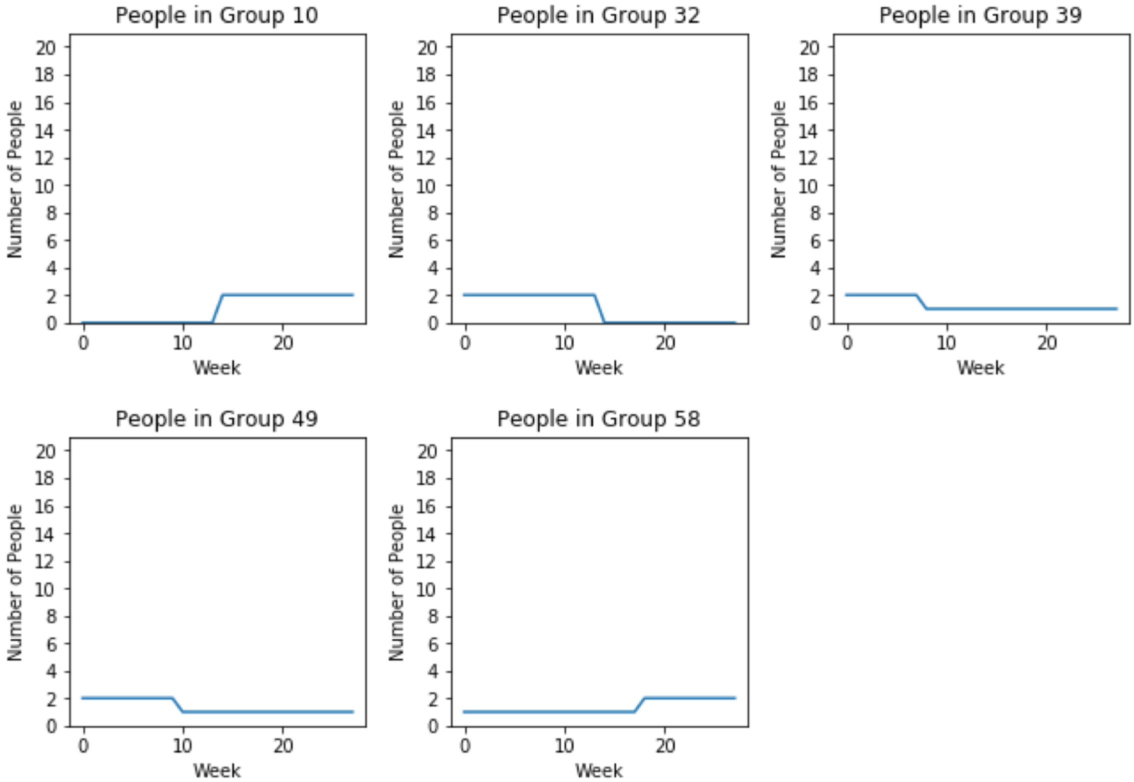

Figure 14. Sizes of interesting groups in Cohort 4 over time. No group is ever larger than 2 people.

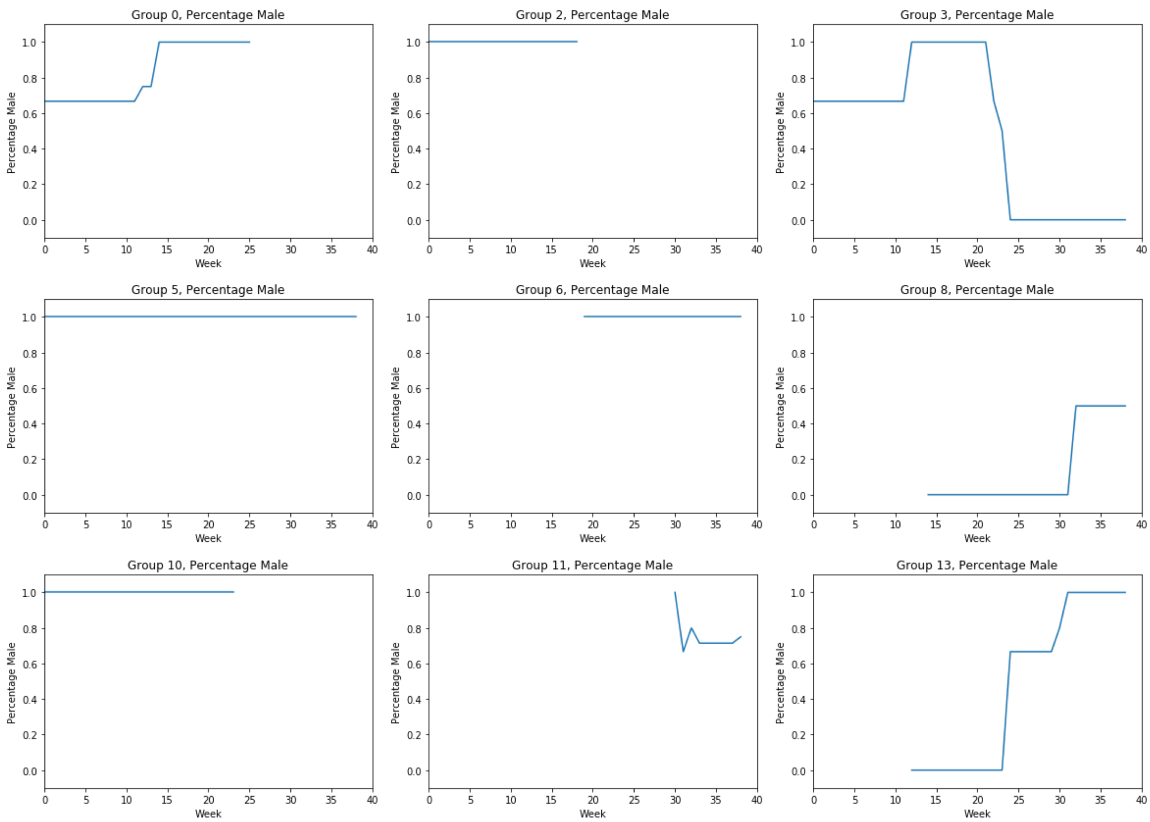

Figure 15. Gender makeup of cohort 1 groups.

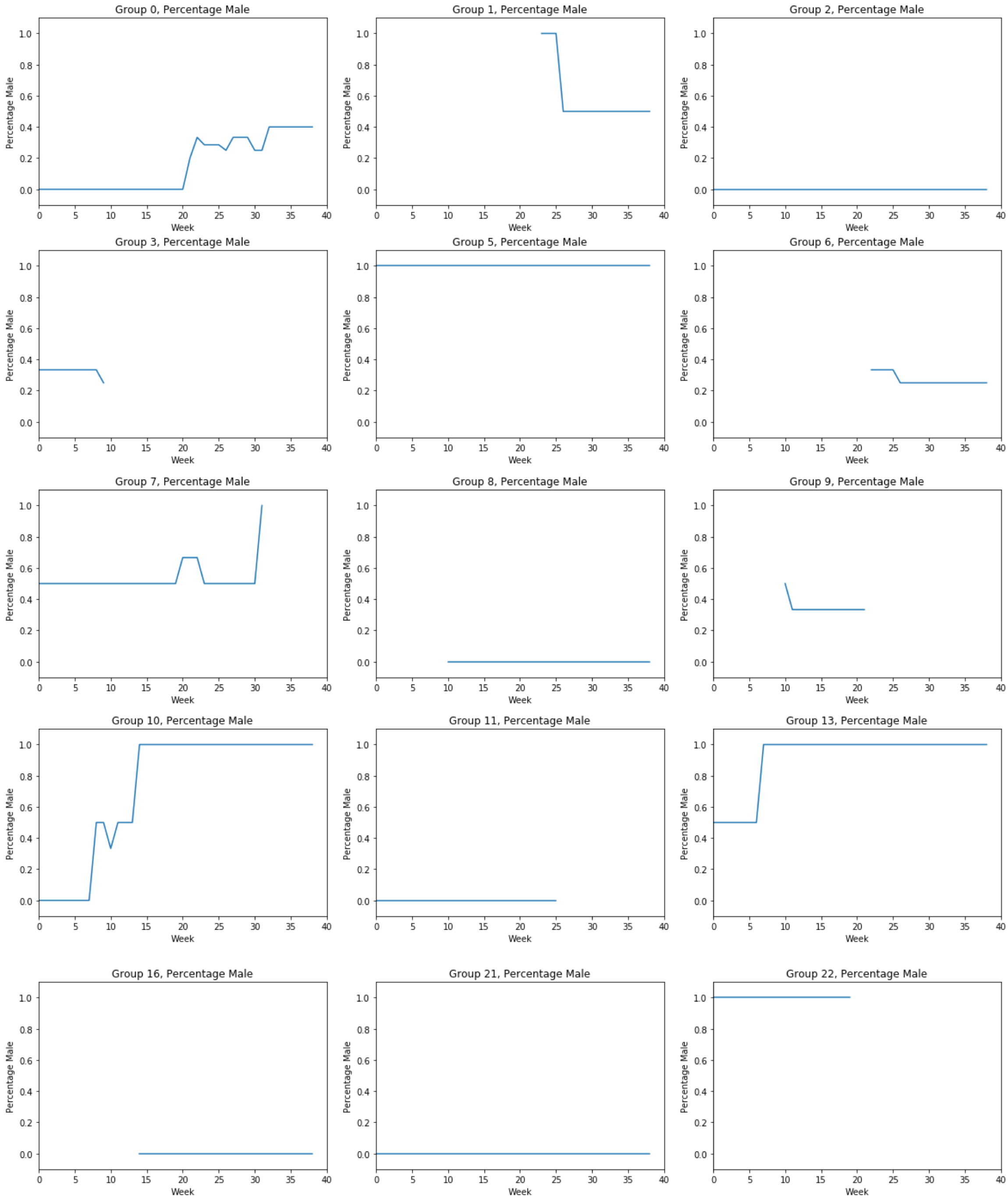

Figure 16. Gender makeup of cohort 2 groups. Group 0 shows interesting behavior.

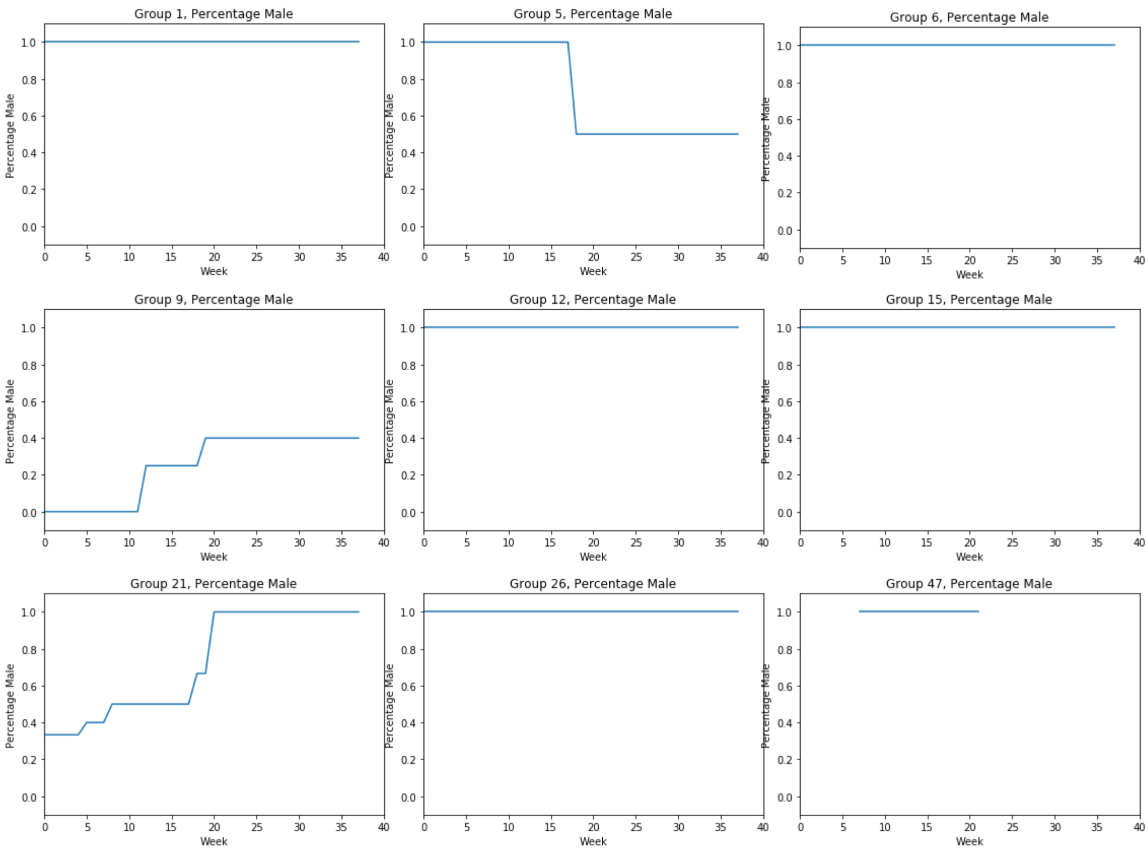

Figure 17. Gender makeup of cohort 3 groups. Groups 9 and 21 show interesting behavior.

Figure 17 shows the gender makeup of interesting groups in cohort 3. Group 9 slightly increases in size of the course of the study while group 21 initially increases but drops to a single person around week 20. This implies that group 21 was initially mixed gender (actually majority female), and then the females leave the group. On the other hand, group 9 is initally all female and then males join the group. Whether this is because of assignments to work groups or organic social behavior cannot be discerned from the data.

We do not consider the gender distribution of groups in cohort 4 due to prohibitively small group sizes.

4.3 Stress levels over time

Finally, we investigate one of the dynamic psychological measures, stress, in combination with the discovered groups. We consider whether each group has a distinct stress level, whether groups tend to diverge in stress level over time, and whether stress level changes in correlation with group size.

Figure 18 shows average stress for each group in cohort 1 where we have sufficient data over time. Unfortunately the large amount of missing data makes conclusions hard to draw.

Figure 19 shows average stress for groups in cohort 2 over time. It appears that smaller groups have more variable stress, although missing data makes this hard to confirm. Follow-up interviews with members of cohort 2 might indicate whether group size correlates with variability of stress level.

Figure 20 shows average stress level for group in cohort 3. Group 1, before compliance ends, has a spike in stress while group 15's stress level steadily decreases. Both groups are all male the entire study, and have relatively stable size.

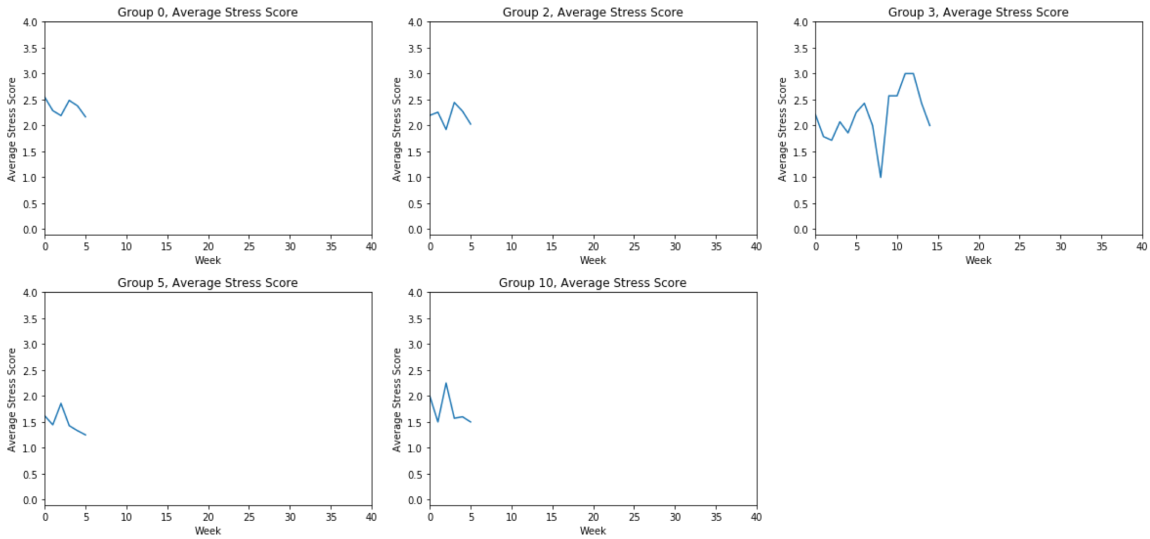

Figure 18. Average stress levels for groups in cohort 1. Groups shown are groups which had enough data to calculate an average. Missing data is prevalent.

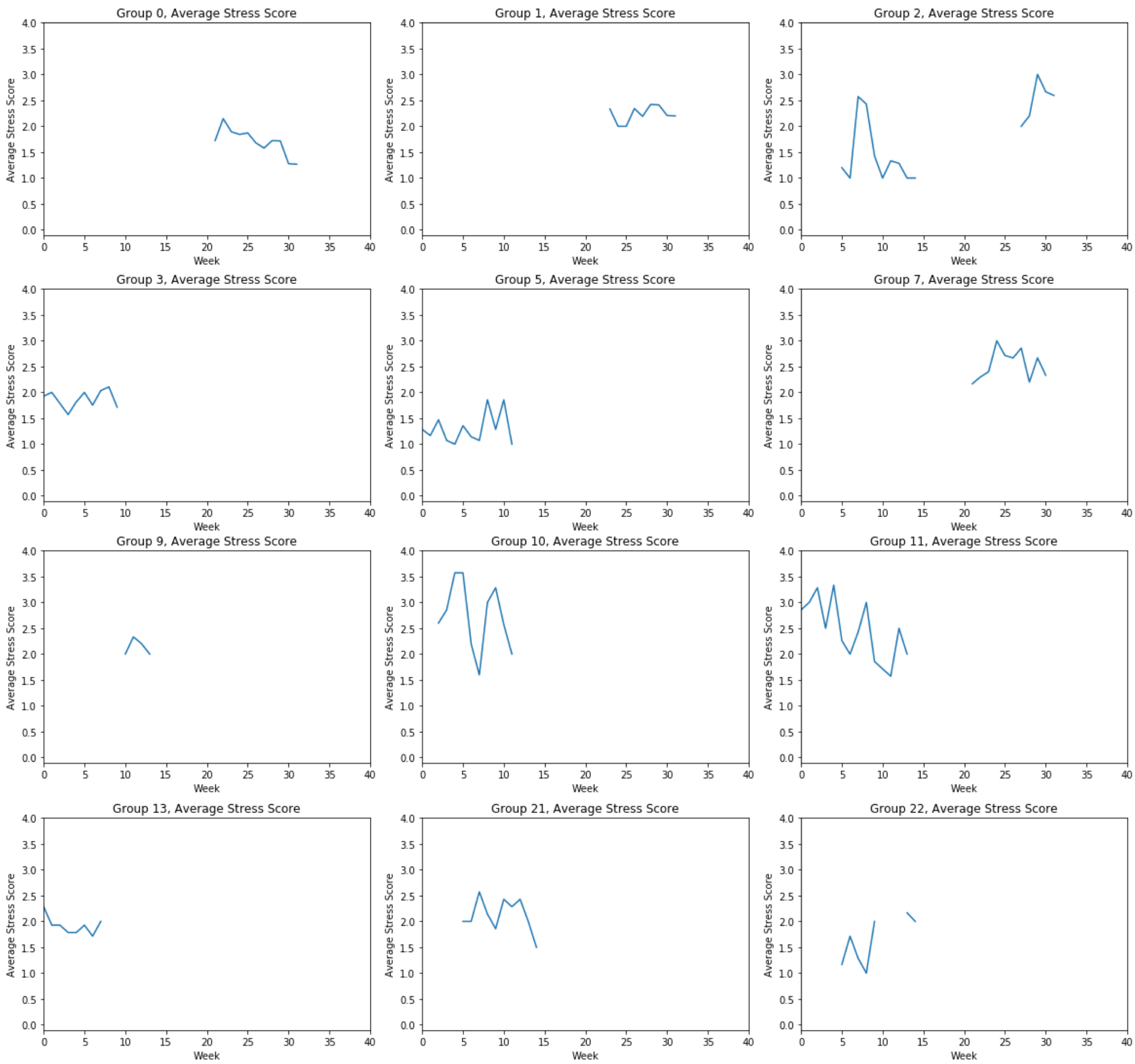

Figure 19. Average stress levels for groups in cohort 2. Smaller groups seem to have more variability.

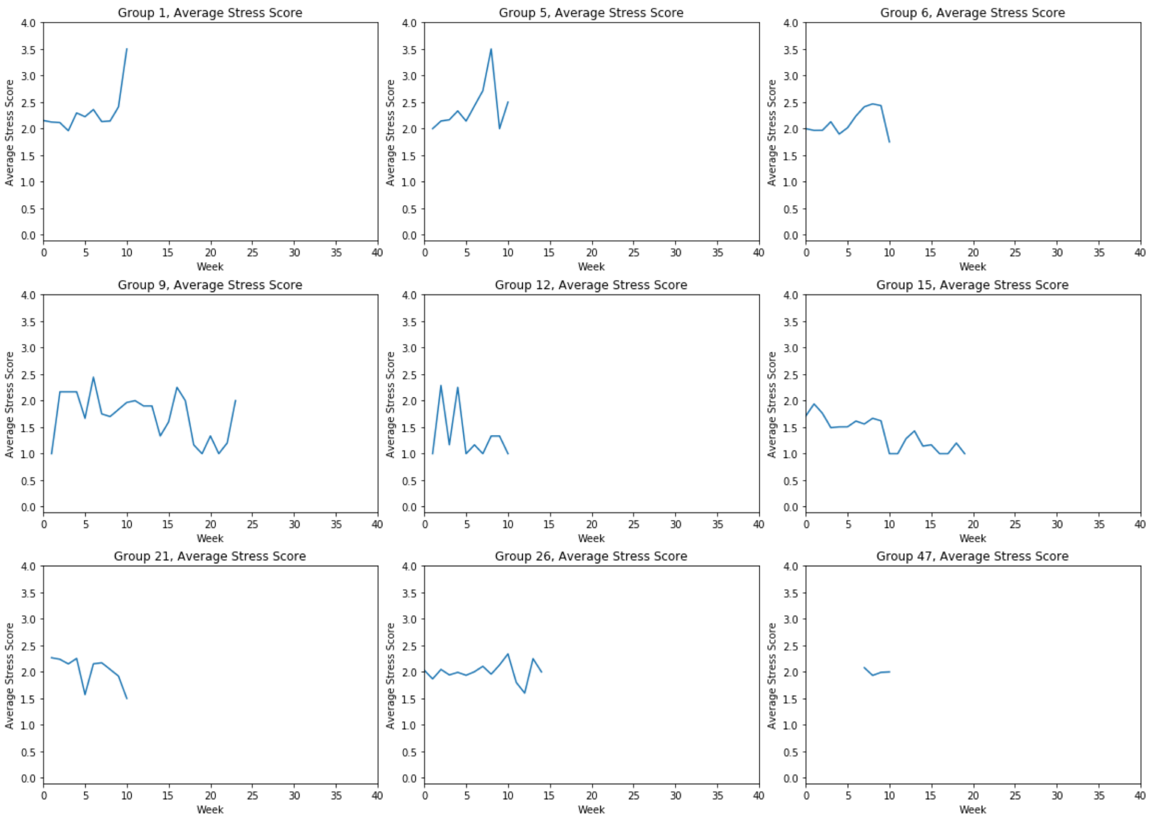

Figure 20. Average stress levels in groups in cohort 3. Participant compliance seems to have fallen off sharply at around week 15.

5 INFLUENCE

Social influence flows through various forms of interaction, including television, social media, professional publications, and face-to-face interactions, for example. Our professional behavior, our social norms, and even our beliefs are all influenced by the social interactions that we have on a daily basis. Here, we attempt to detect a type of *quantitative* social influence that is mediated through direct social interactions (i.e. face-to-face social contact) in the workplace and in non-work environments, to the extent they are represented in the Notre Dame data set. The methods that we use to search for the signal of influence in these data are generalizable to any environment where there are quantitative outcomes of interest and some direct measure of physical interaction.

We consider both a *population-level* and an *interaction-level* influence model. The population-level model conceptualizes influence as a personal characteristic that is homogeneous in time (i.e., a person's influence on other people's measured outcomes is considered in aggregate, rather than on a person-by-person basis).This model has one parameter corresponding to each person in the study that gives the per unit time effect of interacting with that person on the measured outcome of interest. For example, if we are studying the number of drinks a person has per week, and our subject Joe has an estimated influence of 1.5, the model predicts that interacting with Joe for one unit of time will increase the average number of drinks that people have by 1.5 per week. The strength of this model is that it has few parameters and, therefore, has higher power to detect influence. However, because the model effectively assumes that a person's influence is homogeneous (i.e. that it affects all other people equally), it is the most prone to unmeasured confounding from missing social interactions.

The interaction-level model relaxes some of the assumptions of the population model by conceptualizing influence as a property of a dyadic interaction. That is, it considers the different levels of influence a person can have on different people. This prevents some of the confounding issues in the population model, but comes at the price of partial non-identifiability: while the population model has k parameters for k study participants, the interaction model has $\binom{k}{2}$ parameters. Because most pairs of individuals in an institution do not interact or interact very little, we may not have as much data as we need to fit this model, and, in fact, most of the parameters will be completely unconstrained. However, we gain significant explanatory power using the interaction model, as influence is now conceptualized as a feature of a directed interaction between people, which has a more natural interpretation. Also, because we are using a quantitative measure of influence, influence has both a sign (positive or negative) and a magnitude, which can vary between interactions. This approach also allows for asymmetric influence, that is, your effect on me can be different from my effect on you.

5.1 Evidence of social influence

For the population-level model we consider both positive and negative affect measured on a daily to weekly scale. Figure 21 shows the time series of measured negative affect in all of the cohort 1 participants. Negative and positive affect were selected because it is likely that they are influenced by both workplace and social interactions. Figure 21 shows that negative affect is highly variable over time (i.e. that there are features of the data to explain) and that there is not a single dominant secular trend (e.g. all of the employees showing an increasing/decreasing trend in the outcome over the year). Both of the models attempt to explain these patterns in terms of individuals' interactions with others over the study period. We also apply the interaction model to alcohol use, work ethic, and sleep patterns.

Figure 21. Time series of negative affect in cohort 1 participants over the study period

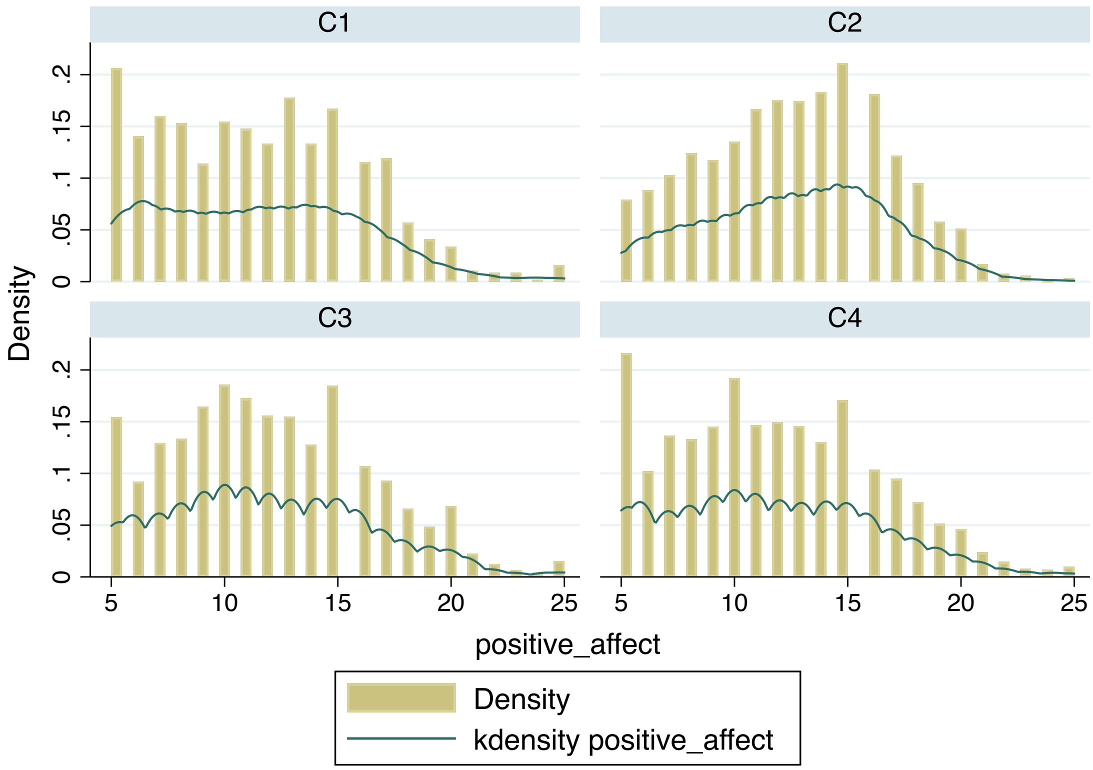

Figure 22. Distribution of the daily value of the variable "positive affect" over cohorts.

Among the many daily variables measuring the psychological well being and the behavior of participants we select metrics related to their emotions to identify mechanisms of social influence. In particular, we want to study if and how social interactions impact the emotional status of participants, and thus we initially focus on the two variables that measure the positive and negative affect of subjects.

Figure 22 presents the distribution of daily values of the positive affect variable across different cohorts and Figure 23 presents the same for negative affect. These figures illustrate a few important points. First, positive affect varies significantly in all cohorts and its standard deviation is hence quite large. Second, negative affect on the contrary is very much concentrated on very low values. Third, the cohorts all look very similar except for cohort 2, which has relatively fewer subjects with low daily values of positive affect.

We count two people as having a face-to-face interaction if their sensors indicate close proximity to one another. This is an imperfect measure as the the sensors do not always agree with one another, that is, reported interactions are not necessarily symmetric. We also do not have any guarantee that a person is always wearing their sensor, which could be removed or left in an office, leading to noisy and biased measurements. As described previously, to account for asymmetry we simply force interactions to be symmetric by taking the maximum interaction time reported by either beacon in a given time period. For example, if beacon A reports an interaction with beacon B of 1 minute and beacon B reports 2 minutes, we assume the two persons were in face-to-face contact for 2 minutes.

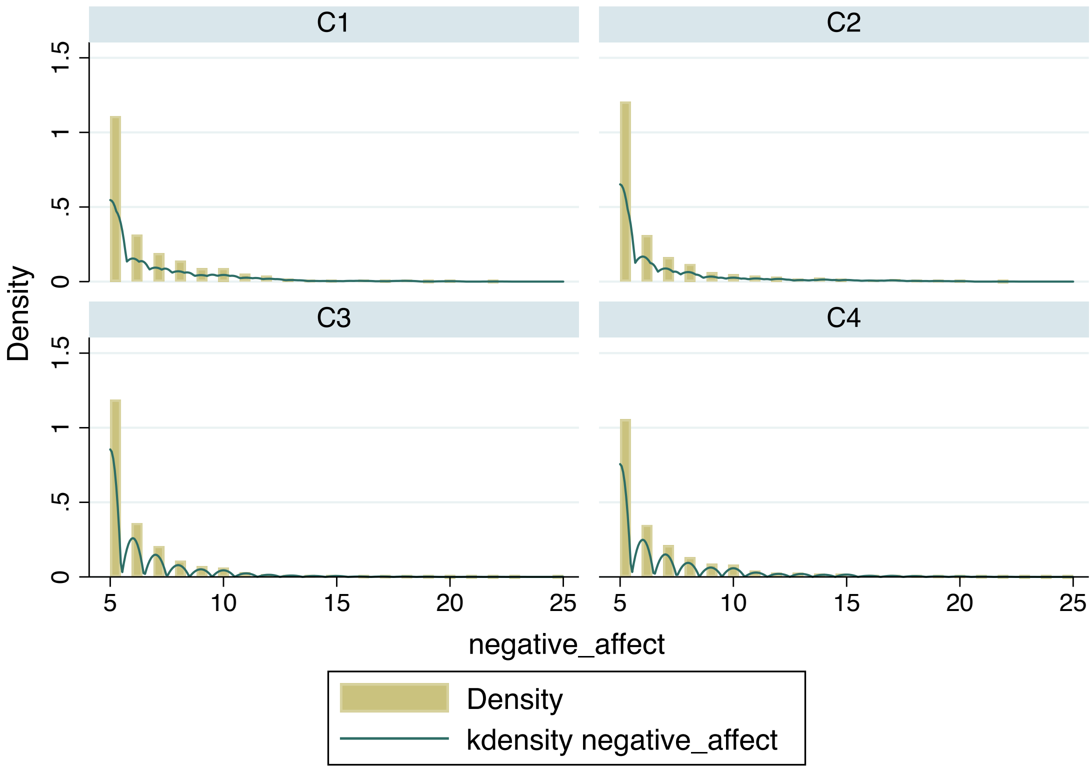

Figure 23. Distribution of the daily value of the variable "negative affect" over cohorts.

5.2 Modeling approaches for identifying social influencers

Several quantitative models have been proposed to describe social influence [see for instance the survey in 13] but only a few approaches have been developed for the analysis of social influence in actual data sets, which are mostly based upon statistical tools [e.g., 11, 1]. We thus analyze the data with a linear regression model, but we also explore the possibility to use a Markov model inspired by the theoretical formal models of social influence developed in the literature [5].

5.2.1 Generalized linear regression models

The generalized linear regression model we use to analyze social influence at the population level when the response is continuous, as it is for positive and negative affect, is as follows:

$$s_i^t = \sum_{j=1}^{N-1} \beta_j d_{j,i}^{t-1} + \epsilon. \tag{1}$$

In the model s_i^t is the status variable (in our case either "positive affect" or "negative affect") of individual i at time (day) t, j is one of the other $N-1$ subjects in the cohort (i.e., $i \neq j$), β_j is the associated regressed coefficient, $d_{j,i}^{t-1}$ is the lagged duration of interaction between subjects i and j, and ϵ is the error term.

The interaction-level social influence model is similar to 1, but models each individual's influence instead of an overall effect. The $\beta_j' s$ from Eq. 1 change to $\beta_{i,j}' s$ but all the other values stay the same as in Eq. 1, as follows:

$$s_i^t = \sum_{j=1}^{N-1} \beta_{i,j} d_{j,i}^{t-1} + \epsilon. \tag{2}$$

$\beta_{i,j}$ is the associated regressed coefficient for person i based upon interactions with individual j.

Equations 1 and 2 are both estimated using a generalized linear model with normal errors. The resulting β_j coefficients should be interpreted as the change of the status variable (i.e., the dependent variable) of all individuals $i = 1, \cdots, n$ due to the interaction with person j the previous day. The $\beta_{i,j}$ coefficients should be interpreted as the change of the status variable (i.e., the dependent variable) of individual i's interaction with individual j the previous day.

Considering the interpretation of βs just described, the analysis can proceed with the identification of influencers as the js having the largest absolute value of β. Influencers are therefore the individuals who are able to change the status of others the most.

The counterproductive work behavior metric is a -1 and 1 response, where -1 is counterproductive and 1 is productive. The binary scale in this metric requires that the data be modeled through a generalized linear model, in particular, logistic regression. The model is similar to 1, except that the response is modeled using a link function, as follows:

$$\log\left(\frac{p}{1-p}\right) = \sum_{j=1}^{N-1} \beta_j d_{j,i}^{t-1} + \epsilon. \tag{3}$$

where $p = P(Y = 1)$ and $1 - p = P(Y = -1)$. In this model, the β_j's are interpreted as the log odds change based on the time that individual i spent with individual j. For this analysis, we select the 10 percent of each cohort with the largest absolute value of β as the influencers we study across multiple survey metrics.

5.2.2 Lasso regression model

Lasso regression is used to shrink some of the interactions between individuals that have spent little time with each other. The lasso is similar to the population-level model, but adds an L1 penalty term in the minimization of the β_js, as follows:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(s_i^t - \sum_{j=1}^{N-1} \beta_j d_{j,i}^{t-1})^2 + \lambda \sum_{l=0}^{p} |\beta_j| \tag{4}$$

where if λ is zero then the equation becomes

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(s_i^t - \sum_{j=1}^{N-1} \beta_j d_{j,i}^{t-1})^2,$$

which is regular OLS Eq. 1. However, if λ is nonzero then some of the β_j values are driven to zero. The larger the value of lambda the more features are shrunk to zero. This can eliminate some features entirely and give us a subset of predictors that helps mitigate multi-collinearity and model complexity especially when the data are sparse. Any β_js that are not zero signify that they have much more weight, or in the case of social influence, that they are the individuals that have the most influence.

5.2.3 Markov model

We also considered an alternative formalism to model influence that does not directly include measured interactions, but rather looks for evidence of potential influence directly in the distribution of outcomes over time in the whole population. Early studies of social influence [e.g., 12, 6, 7] modeled social influence mechanisms as probabilistic changes to some internal state of individuals due exogenous forces such as social interactions. The structure of those models thus resemble Markov chains, and this feature opens the possibility to use their formalization as Markov models to analyze social influence.

In particular, we can consider a Markov model as in:

$$s_i^t = M_i D^{t-1}, \tag{5}$$

where the state s_i^t of individual i at time (day) t depends on a Markov transition matrix M_i and the lagged duration of interaction with each other participant in the cohort D^{t-1}.

The matrix M_i hence contain the state transition probabilities for subject i and depending on interactions with other participants. Since the M matrix can be derived as distributions of data by observing how states changed in each individual after interactions, it is possible to interpret transition probabilities as the impact on i status of interacting with other subjects j. Similarly as before, subjects that have on average the largest chance of changing the state of others can be identified as social influencers. We thus average transition probabilities for each j across is, and we select as influencers the 10 percent with the highest probability of changing others' states.

5.3 Results

5.3.1 Population-level influence

The position of influencers in the whole network, as inferred by the linear model and Markov models, are shown in figures 24 and 25 respectively. The models did not necessarily agree on who was the most influential (a person identified as an influencer on positive affect by the linear regression model only had a 24% chance of being identified as an influencer with the Markov model). However, this is expected as the Markov model was attempting to discern influence without using the underlying interaction network, and was thus considering possible influence between people that never interacted. Nevertheless, the Markov model offers an alternative and confirmatory way to measure influence when there is little or no interaction data present. Because our underling assumption is that the kind of influence we are looking at is mediated through direct social interactions, we focus on the linear regression results that take the empirical interaction times into account.

Figures 26, 27, 28 and 29 compare centrality measures across cohorts for influencers (based on negative and positive affect) and others, as identified by the linear regression model identified as in Eq. 1. Node-level statistics reported in those figures are, starting from the top left graph in each figure, betweenness centrality, closeness centrality, eigenvector centrality, farness score, Katz centrality, nearness score, and strength (i.e., degree centrality). Closeness centrality, farness score and nearness score are closely related, since nearness is the reciprocal of farness, and closeness is equal to the product of nearness and the number of nodes minus 1.

Results shown in Figures 26, 27, 28 and 29 indicate that influencers are often less central than others. This conclusion is supported by measures of betweenness centrality, eigenvector centrality, and strength (degree centrality) in cohorts 1 and 2. Further results supporting this conclusion are the degree centrality of

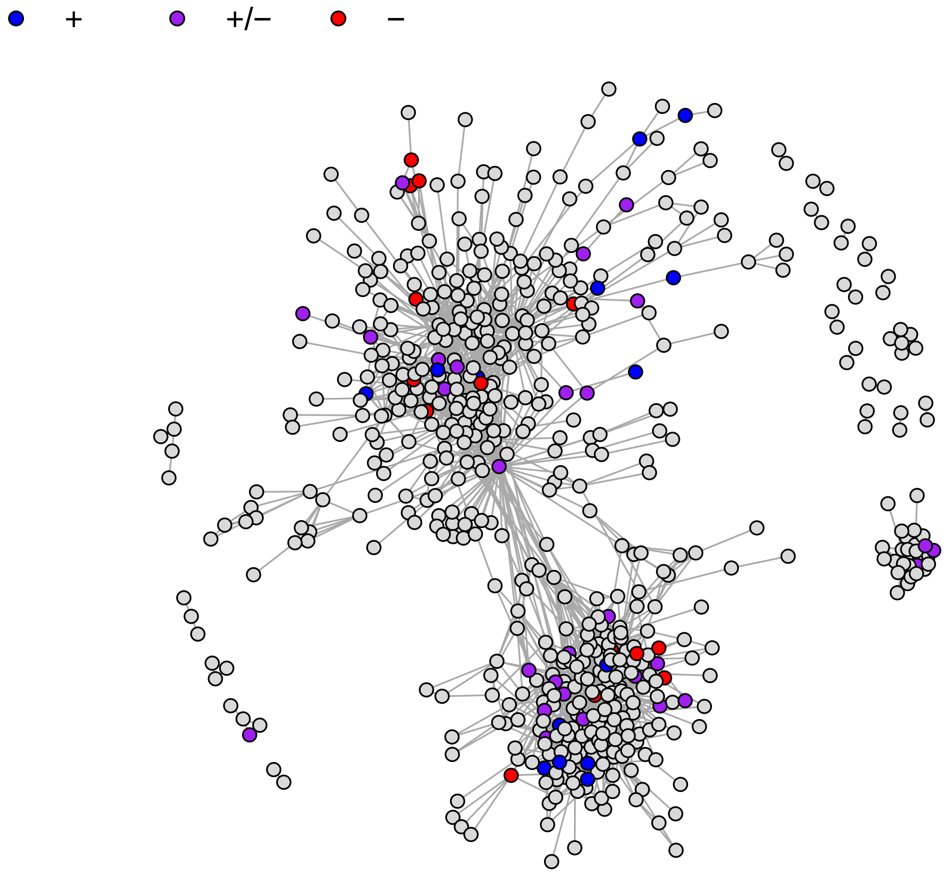

Figure 24. Position of influencers for positive and negative affect in the whole network for the population linear regression model. Influencers of postiive affect are shown in blue while influencers of negative affect are shown in red. Persons who were influential in both measures are shown in purple

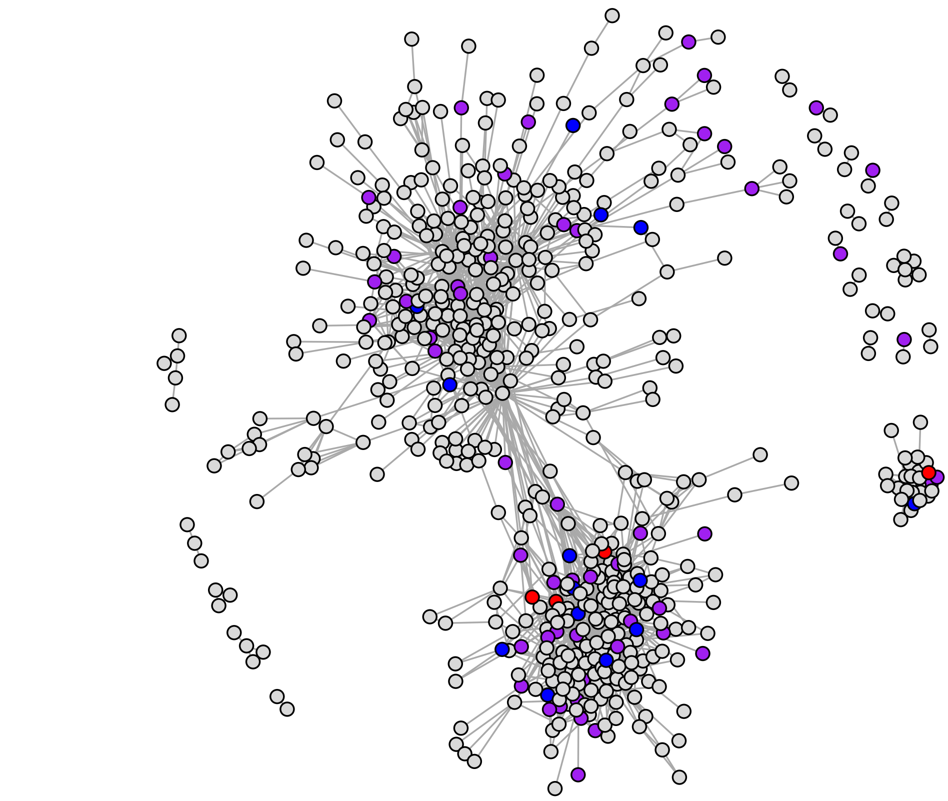

Figure 25. Position of influencers for positive and negative affect in the whole network for the population Markov model. Influencers of postiive affect are shown in blue while influencers of negative affect are shown in red. Persons who were influential in both measures are shown in purple

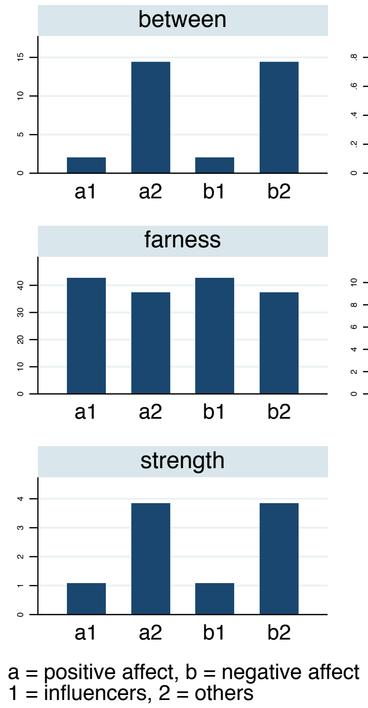

Figure 26. Centrality measures of influencers and others in cohort 1 (linear regression model).

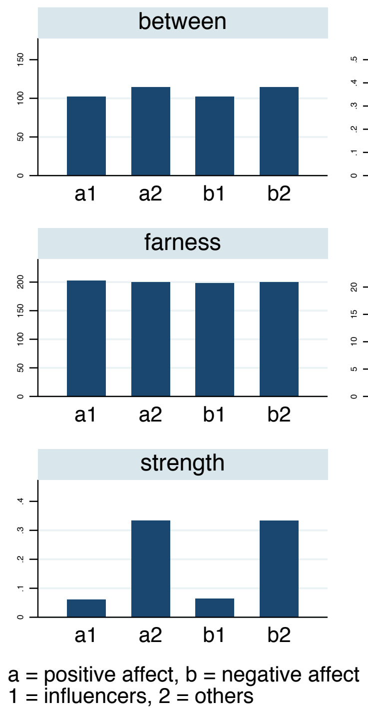

Figure 27. Centrality measures of influencers and others in cohort 2 (linear regression model).

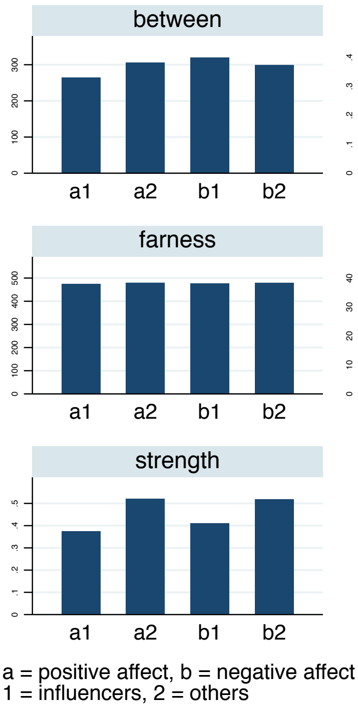

Figure 28. Centrality measures of influencers and others in cohort 3 (linear regression model).

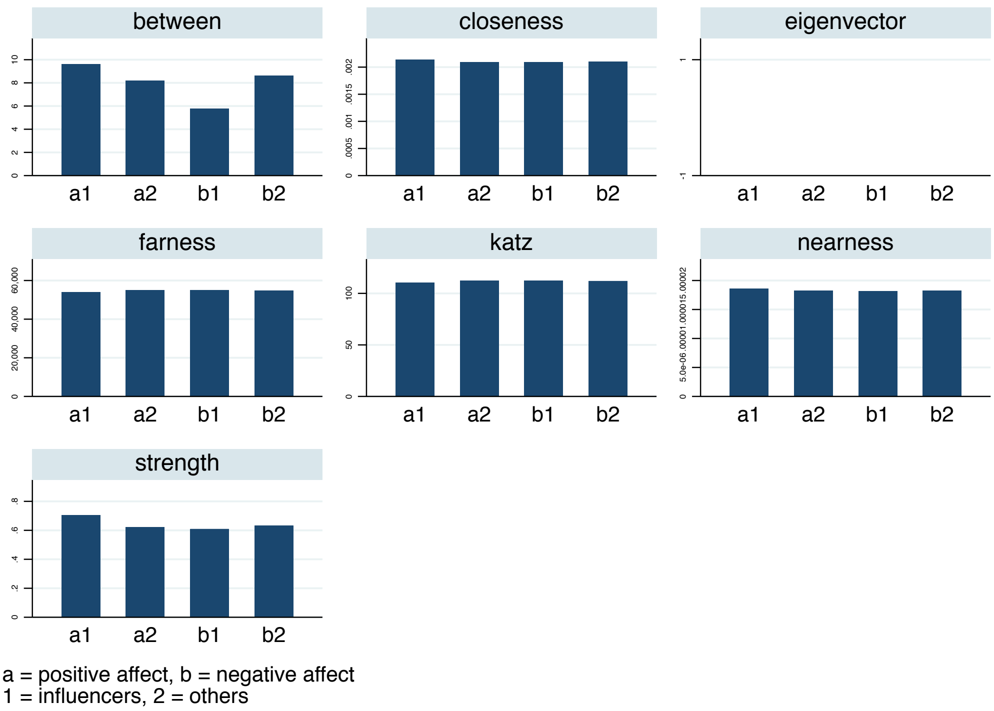

Figure 29. Centrality measures of influencers and others in cohort 4 (linear regression model).

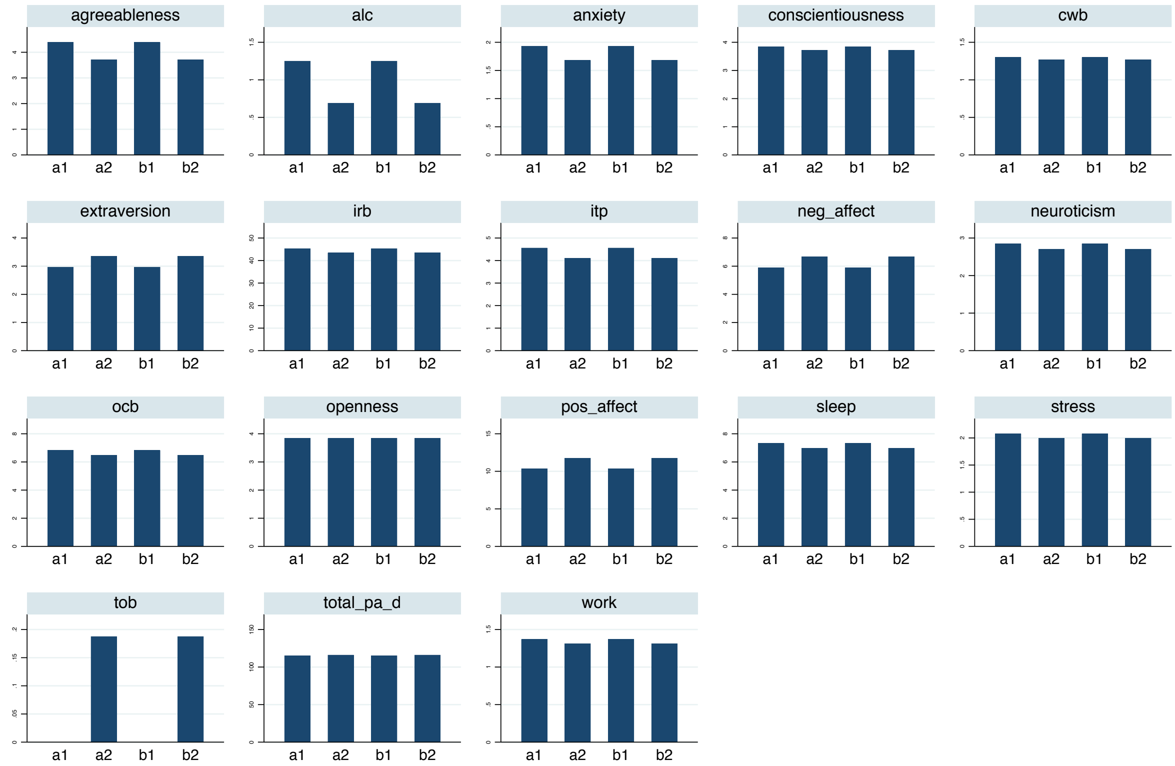

Figure 30. Average scores of influencers and others in cohort 1 (linear regression model).

influencers in cohort 3 and the betweenness centrality in the same cohort, but only for the positive affect model. Other measures and cohort values vary significantly and do not provide firm evidence for or against this finding.

Figures 30, 31, 32, and 33 present average scores of influencers and others as identified in the linear regression model of Eq. 1 for the negative and positive affect metrics. Each Figure presents results for a different cohort. Scores are averaged not only within the two categories of study participants we consider (i.e., influencers and others) and for each affect variable (i.e., positive and negative affect) but also within the period of observation independently from the frequency of observation.

There are only a few consistent results across the four cohorts considered. First, there are only a few tobacco users in all the cohorts who belong to the influencers category. Second, it looks like influencers consume more alcohol than others, although the opposite seems to be true in cohort 4. Third, negative affect is usually slightly lower in influencers (with the only exception of the positive affect model of cohort 4). Similarly, positive affect is lower for influencers, with the exception of negative affect models of cohorts 3 and 4.

With less consistency, in particular due to the results of cohort 4, a few more patterns emerge. Because of diminished consistency they should be considered less robust than those just pointed out. Among them, we notice in influencers a slightly larger value of the two positively correlated variables anxiety and stress. Second, among the big five personality traits we notice in some cohorts that influences have smaller values

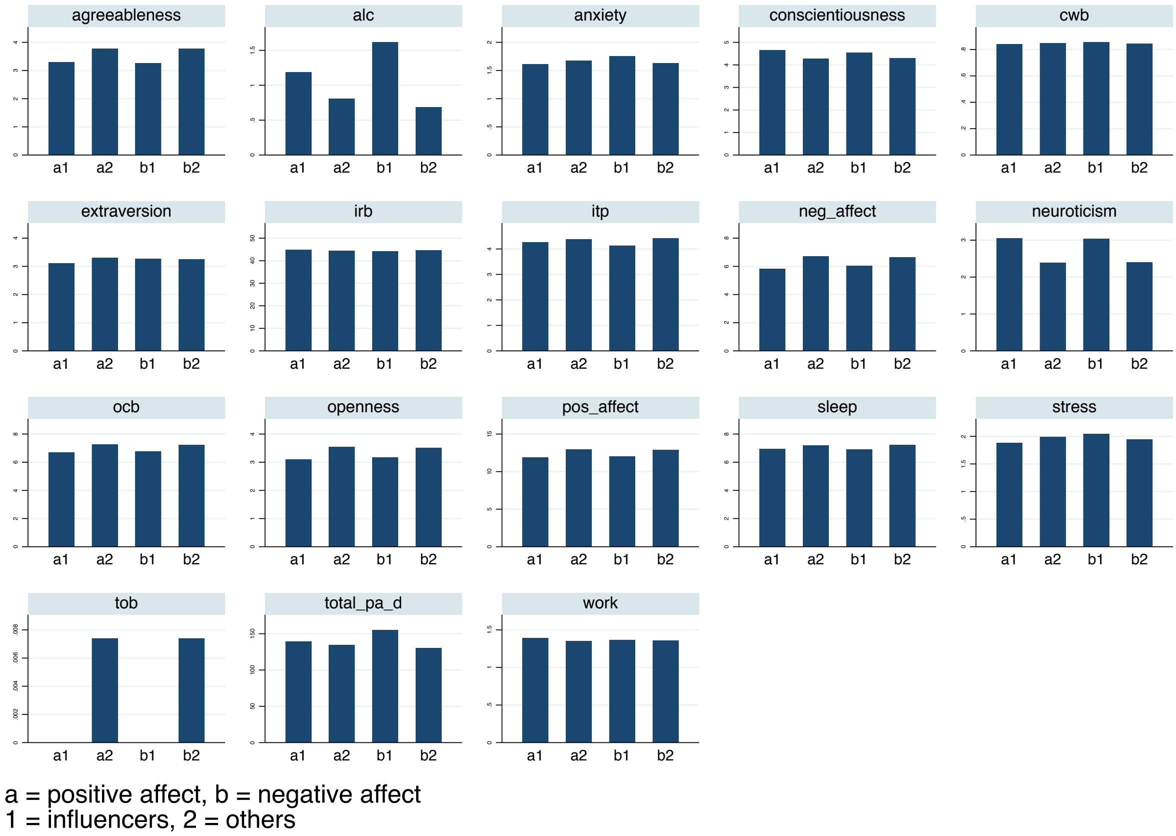

Figure 31. Average scores of influencers and others in cohort 2 (linear regression model).

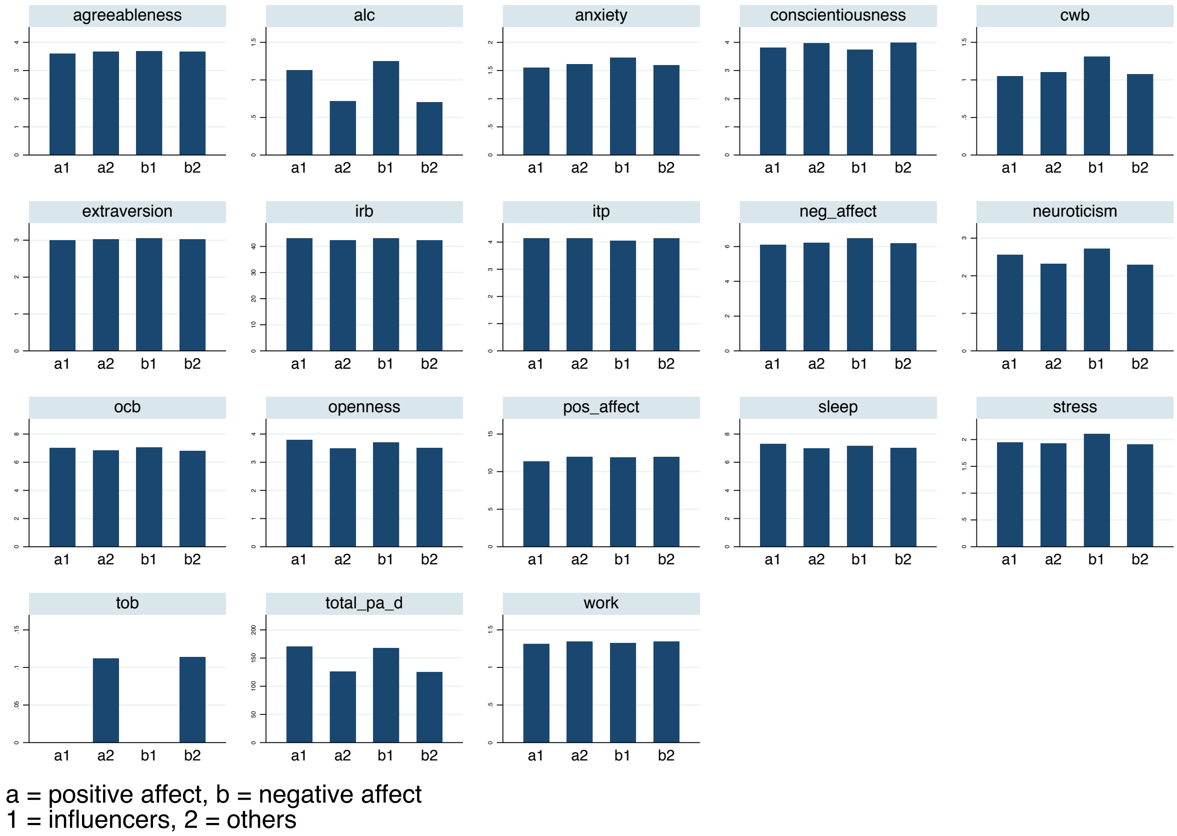

Figure 32. Average scores of influencers and others in cohort 3 (linear regression model).

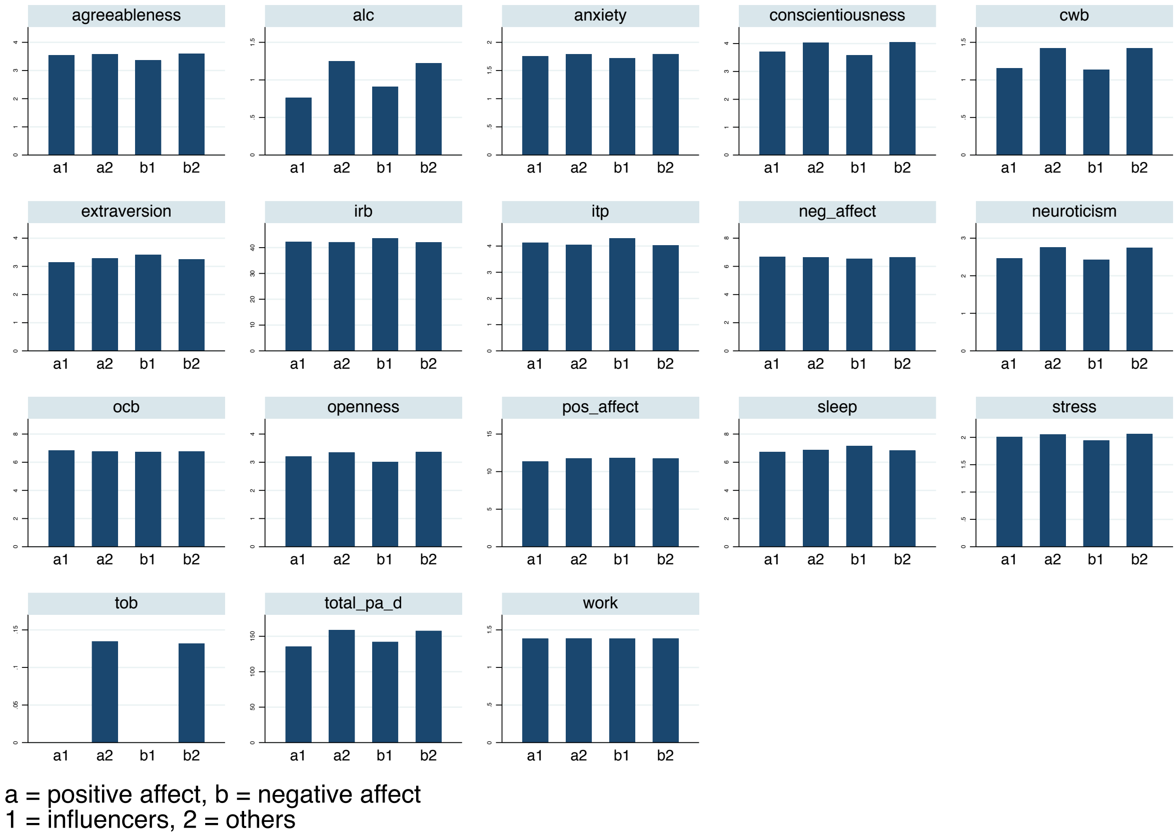

Figure 33. Average scores of influencers and others in cohort 4 (linear regression model).

of extraversion and larger values of neuroticism. All the other variables and personality factors either present no difference between influencers and others or show differences that vary too much across cohorts to be considered robust.

5.3.2 Population-level influence using lasso regression

The second population-level social influence model used lasso regression to see if there were influencers that were consistent across the following survey metrics: positive affect, negative affect, alcohol use, hours of sleep, and counterproductive work behavior. All the survey metrics were on a continuous scale except for counterproductive work behavior, which is on a binary scale.

The model fit results for cohort 1 are shown in Figure 34. In general, it appears that individuals that have some sort of influence tend to be influential across several metrics. For example, two individuals are consistently positive influencers, whereas two individuals are consistently negative influencers to some degree. The magnitude of the negative influences is much larger than the positive influences, which is an interesting result. Overall, the negative influencers decreased positive affect scores the most. Notably, although we did not test this quantitatively, the influential individuals do not appear to be managers, according to the demographic data were were provided for cohort 1, and managers do not appear to be consistently more influential than others in any of our influence models. This is a finding that could be investigated further using the data we already have.

The β_j fits for cohort 2 are shown in Figures 35 and 36. In this cohort, there are three positive influencers and one negative influencer. Positive affect was the scale that most positively influenced individuals in the cohort, with a significant score increase. The counterproductive work behavior metric was the value that decreased the most, meaning that spending time with individual C2-18 decreased the log odds of having a counterproductive work behavior.

The β_j fits for cohort 3 are shown in Figures 37, 38, 39, and 40. There were three individuals that overall positively influenced others' survey scores. Positive and negative affect were the metrics that were the most influenced with the β coefficients much higher than the other metrics. There were also three individuals that had a negative influence on the others in the cohort. The overall positive affect score decreased significantly based upon interactions with the negative influencers.

The β_j fits for cohort 4 are shown in Figures 41, 42, 43, and 44. In this cohort, all the influencers had a positive impact on survey scores. This is likely due to only a few individuals having interactions with each other as seen from the network models in Figures 79-81.

Overall, the individuals identified as influencers are consistent across multiple metrics for all the cohorts. There are more individuals that are likely to positively impact scores compared to those who negatively impact scores. The exception is in cohort 1, the smallest cohort. This could be due more close connections, which is opposite of cohort 4, which has fewer connections.

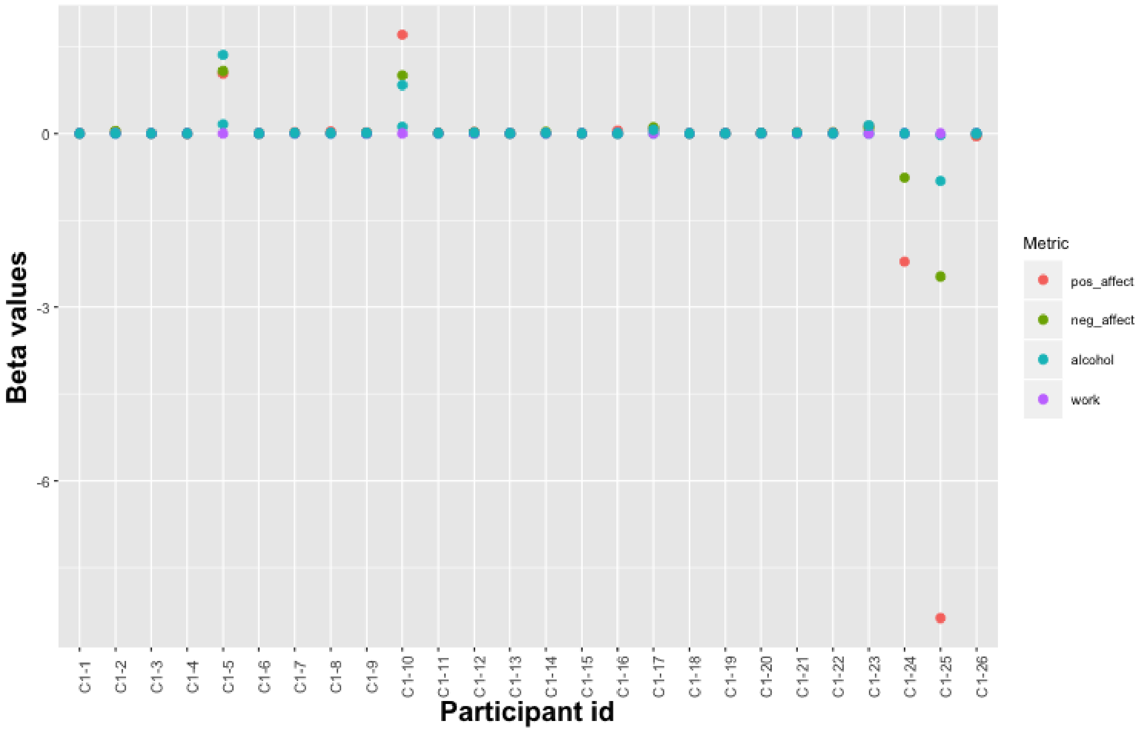

Figure 34. Model fits, β_js for cohort 1 for positive affect, negative affect, sleep and work for all unblinded participants (lasso regression model).

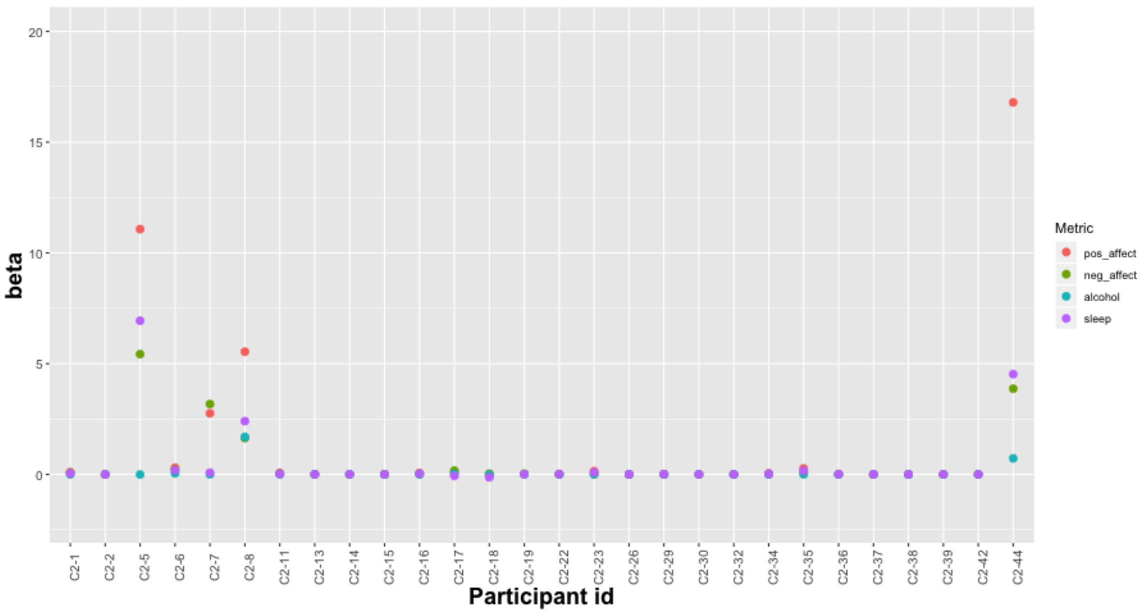

Figure 35. Model fits, β_js for cohort 2 for positive affect, negative affect, sleep and work for unblinded participants 1-44 (lasso regression model).

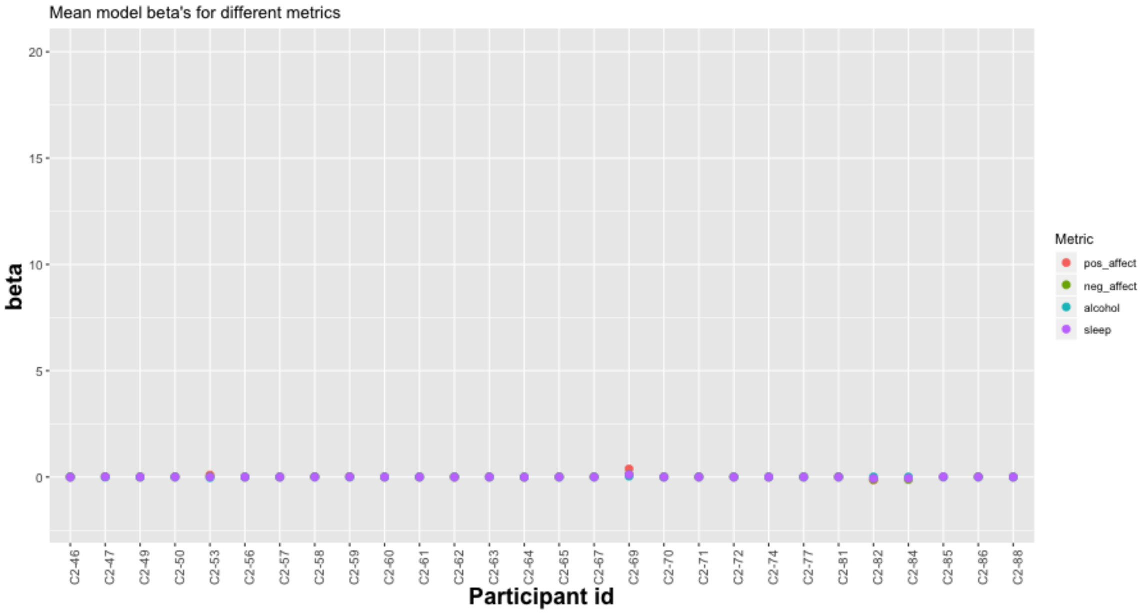

Figure 36. Model fits, β_js for cohort 2 for positive affect, negative affect, sleep and work for unblinded participants 46-88 (lasso regression model).

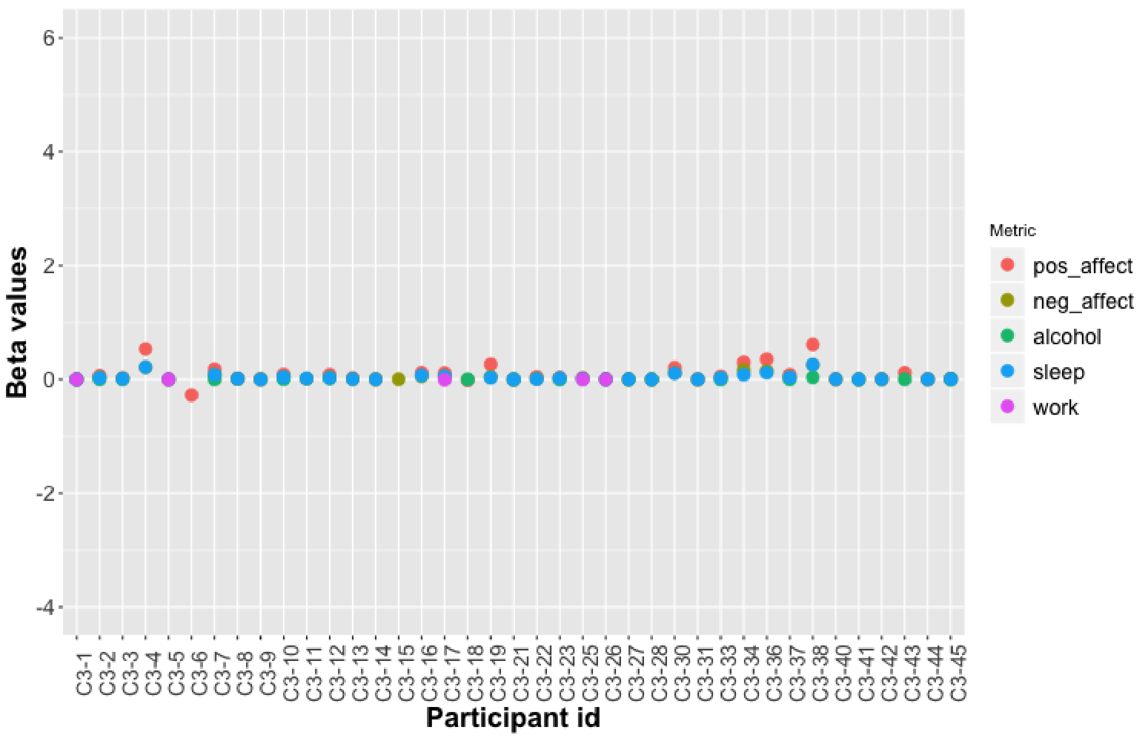

Figure 37. Model fits, β_js for cohort 3 for positive affect, negative affect, sleep and work for unblinded participants 1-45 (lasso regression model).

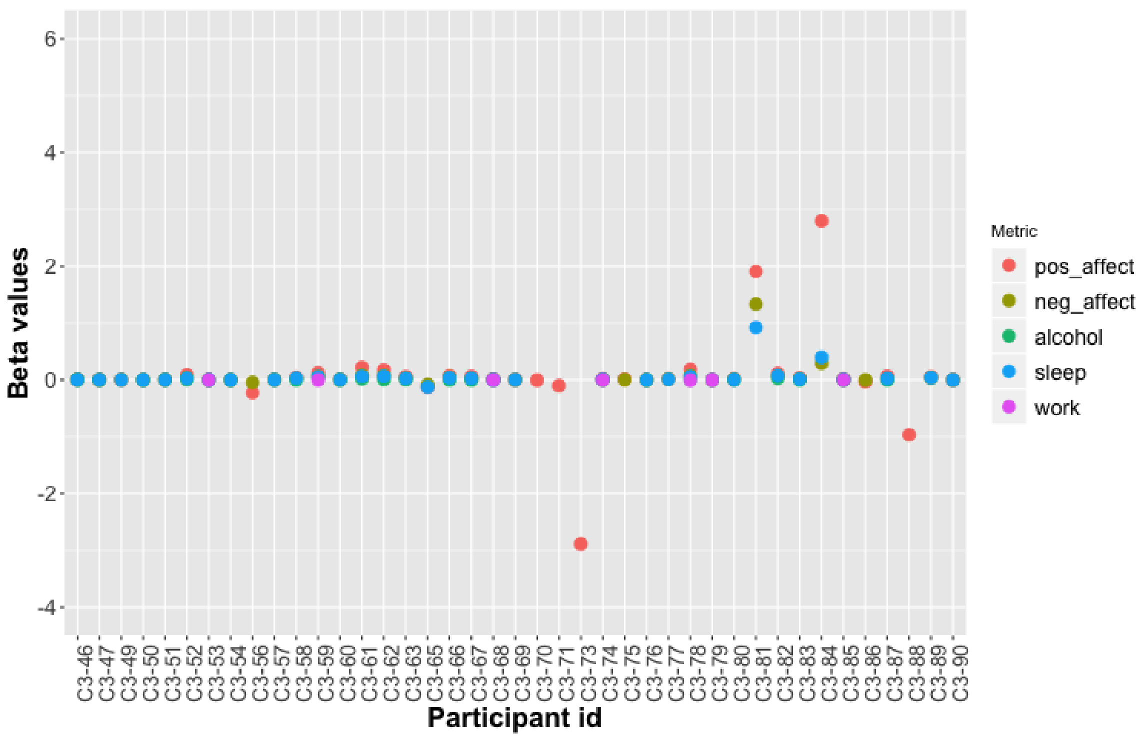

Figure 38. Model fits, β_js for cohort 3 for positive affect, negative affect, sleep and work for unblinded participants 46-90 (lasso regression model).

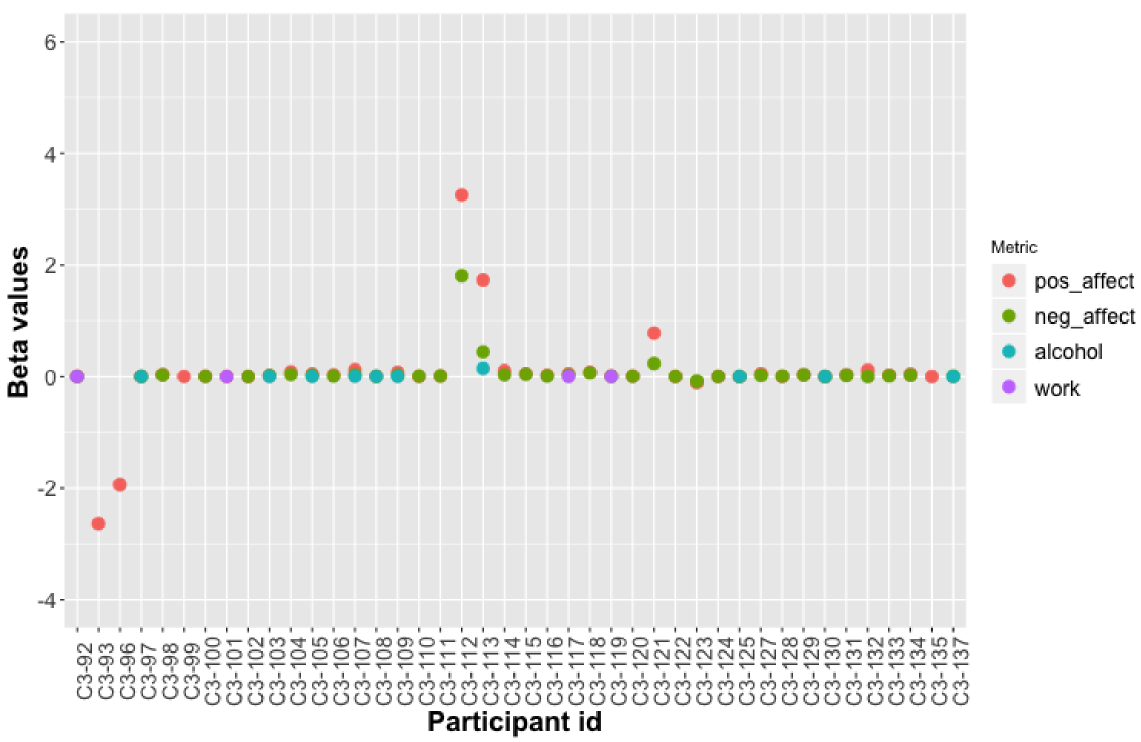

Figure 39. Model fits, β_js for cohort 3 for positive affect, negative affect, sleep and work for unblinded participants 92-137 (lasso regression model).

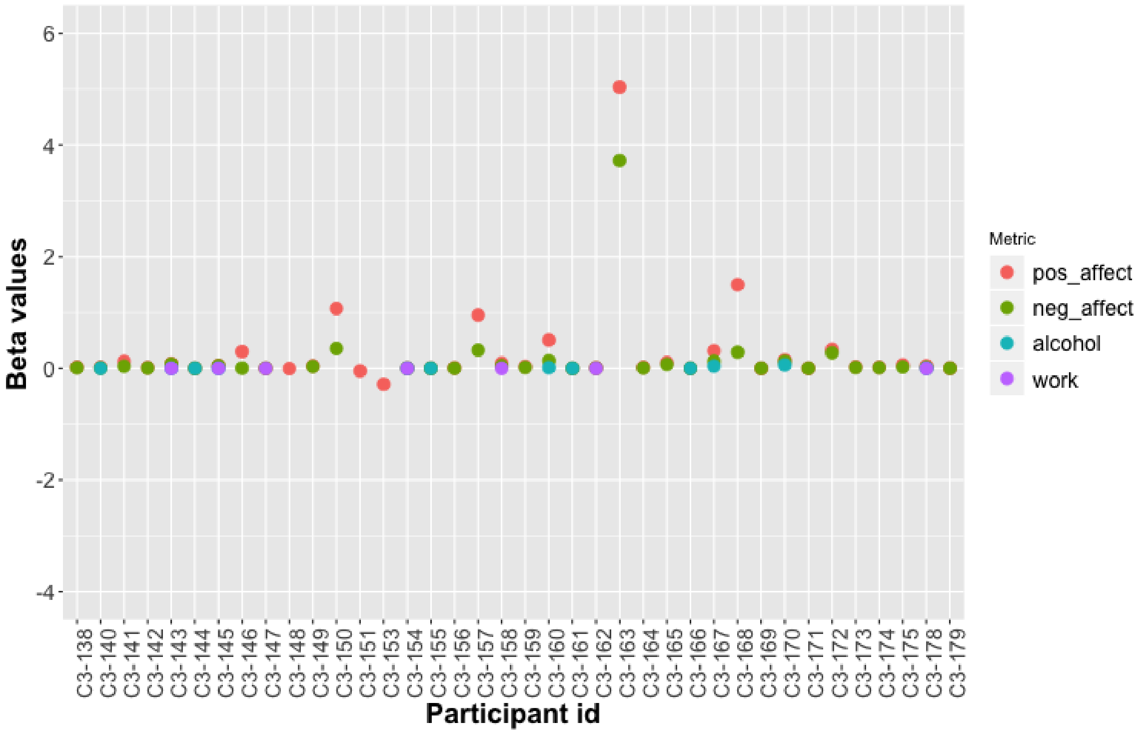

Figure 40. Model fits, β_js for cohort 3 for positive affect, negative affect, sleep and work for unblinded participants 138-179 (lasso regression model).

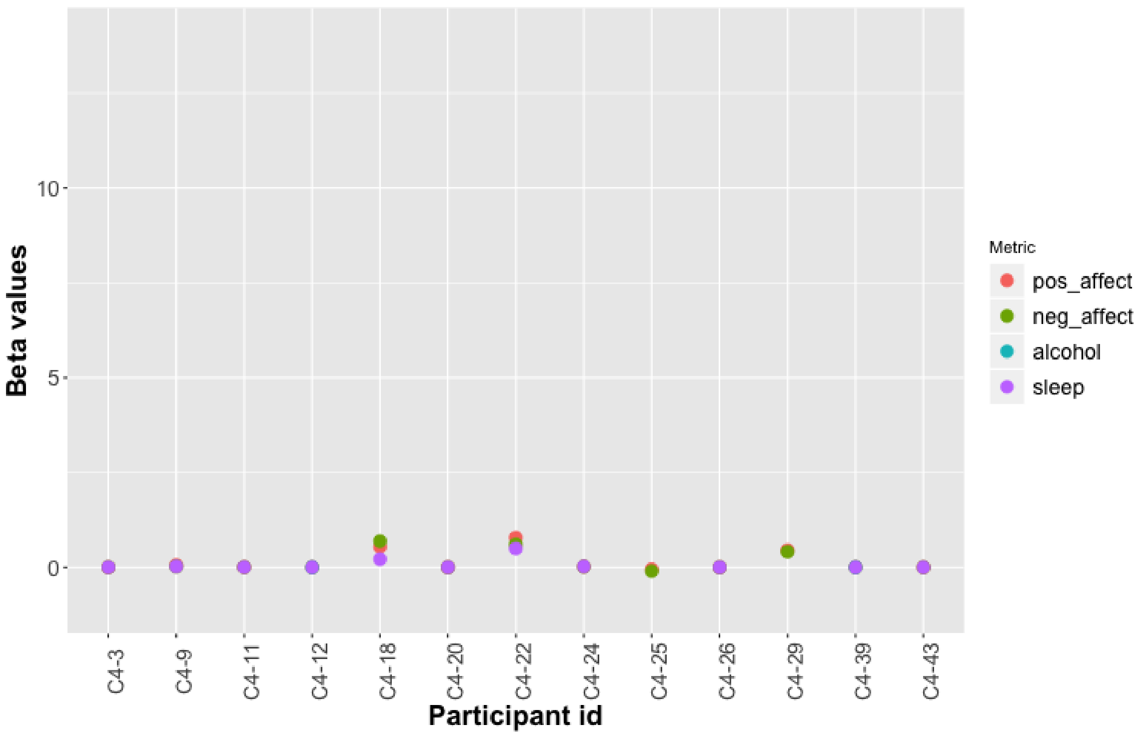

Figure 41. Model fits, β_js for cohort 4 for positive affect, negative affect, sleep and work for unblinded participants 1-43 (lasso regression model).

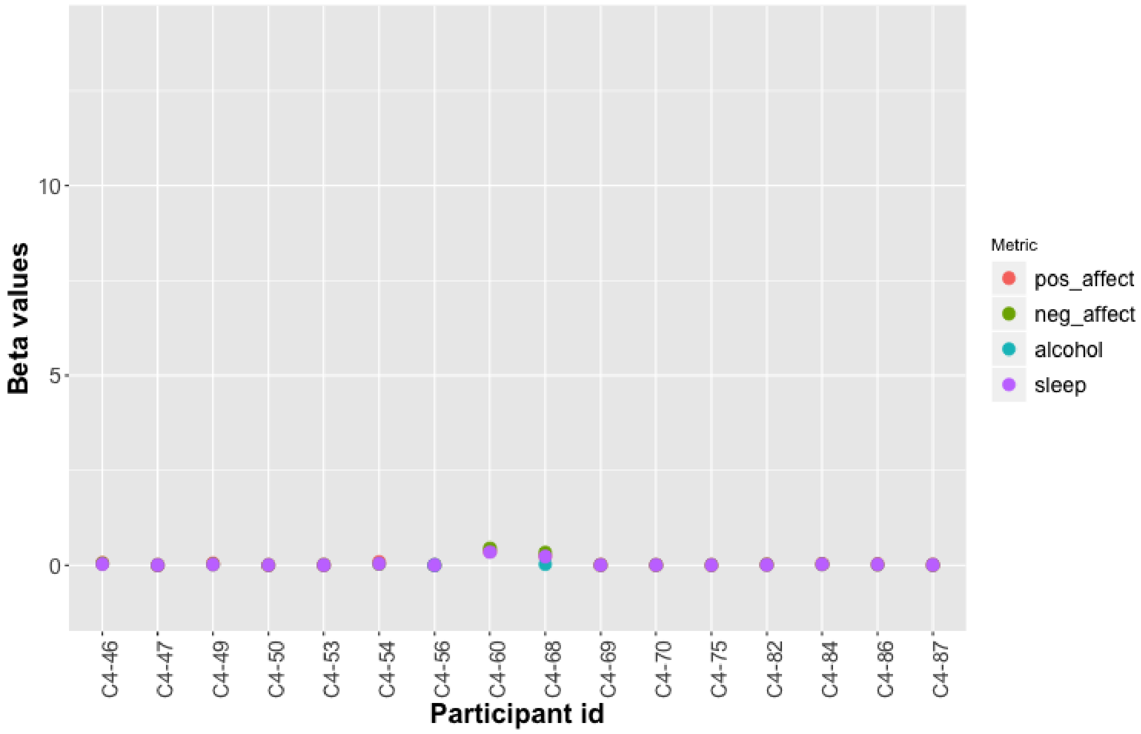

Figure 42. Model fits, β_js for cohort 4 for positive affect, negative affect, sleep and work for unblinded participants 46-87 (lasso regression model).

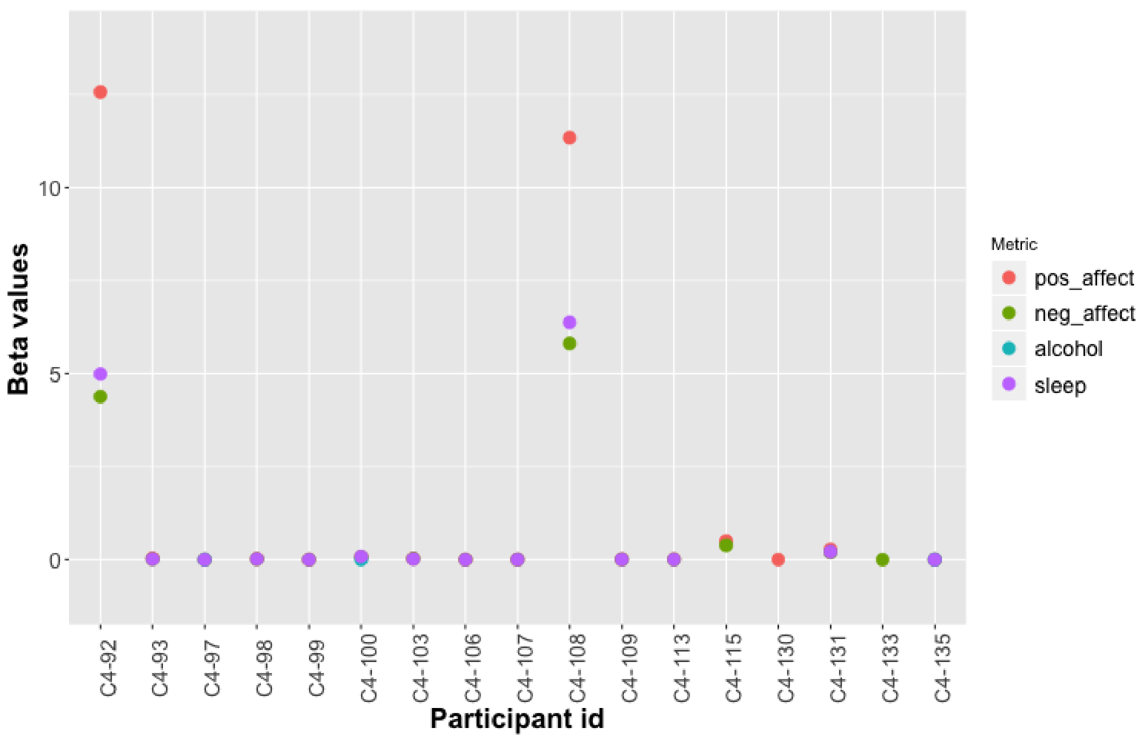

Figure 43. Model fits, β_js for cohort 4 for positive affect, negative affect, sleep and work for unblinded participants 92-135 (lasso regression model).

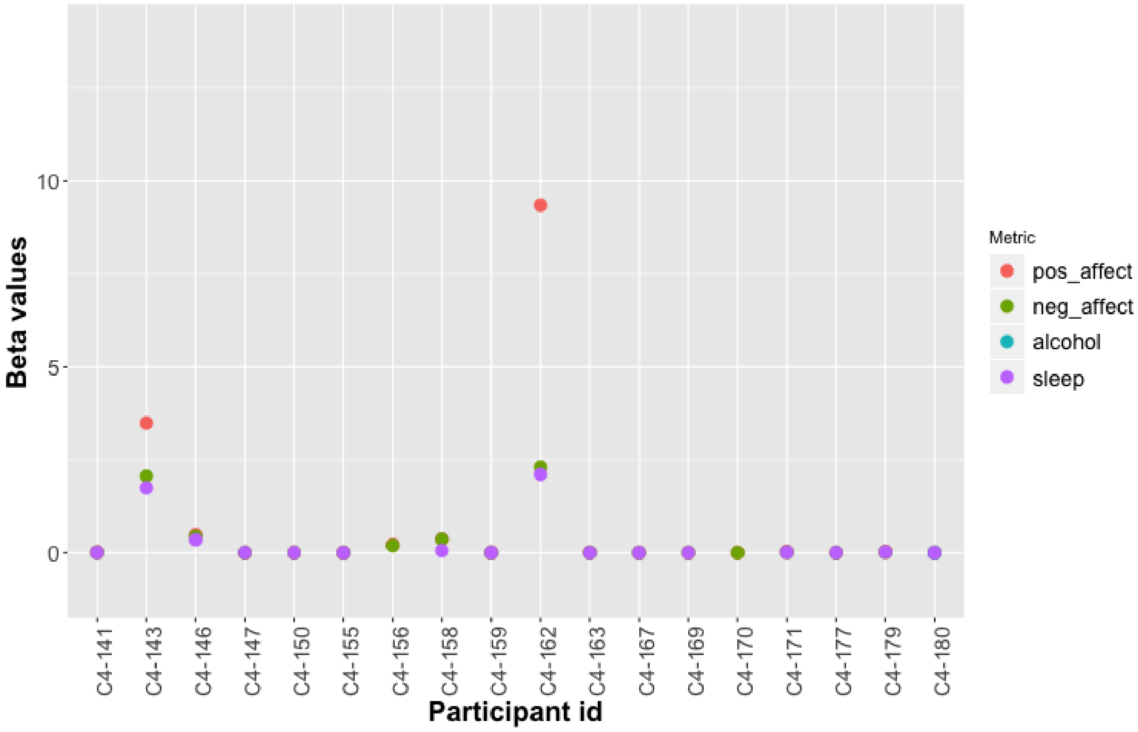

Figure 44. Model fits, β_j's for cohort 4 for positive affect, negative affect, sleep and work for unblinded participants 141-180 (lasso regression model).

5.3.3 Interaction-level influence

The interaction-level model looks at the time individual i spent with individual j, where $i \neq j$, across all possible connected pairs. This model is used to help represent individual effects vs. the population effect. We constructed network graphs to reflect the magnitude of the individual predictor fits for the following survey metrics: positive affect, negative affect, alcohol use, and hours of sleep.

The network graphs of the predictor fits for cohort 1 are shown in Figures 45, 46, 47, and 48. In cohort 1, there are only a few individuals who appear to have a large role in people's responses. For example, the positive affect influence model has three individuals that contribute to an individual having a higher positive affect score, and three individuals that contribute to lower positive affect scores. In general, cohort 1 only has a few individuals who seem to influence survey scores, between 1-3 people positively and 1-3 people negatively.

The network graphs of the predictor fits for cohort 2 are shown in Figures 71, 72, 73, and 74 in Appendix B. In cohort 2, there are generally more individuals that positively influence an individual's survey score as denoted by the stronger blue lines. For example, the positive affect influence model has approximately seven individuals that contribute to a higher value of positive affect and only two main individuals that contribute negatively.

The network graph of the predictor fits for cohort 3 are shown in Figures 75,76, 77, and 78 in Appendix B. Cohort 3 has over one hundred individuals, making it difficult to see the specific individuals that have a positive or negative influence on others. The main takeaway here is that there are more positive connections (blue lines) than negative connections (red lines). Additionally, the people in cohort 3 seem to be fairly well connected, implying that individuals interact with several other people throughout the day.

The network graph of the predictor fits for cohort 4 are shown in Figures 79, 80, and 81 in Appendix B. Cohort 4 appears to be much different than the others. Although the size of the cohort is similar to cohort 3, the connections are more sparse. This means that people do not spend as much time with others according to the duration time used in the model.

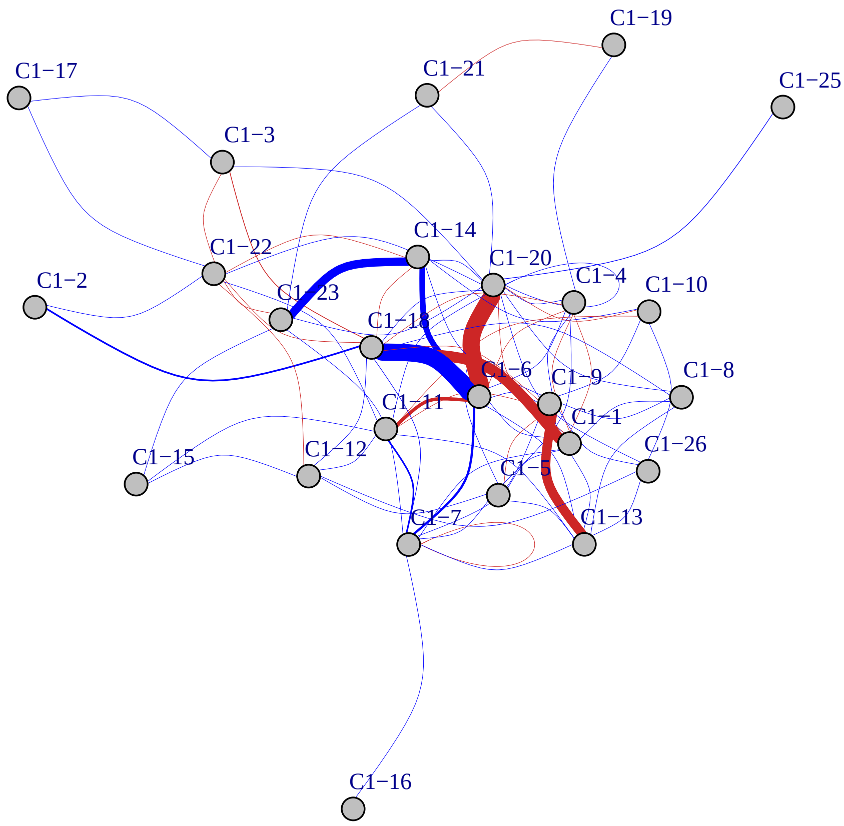

Figure 45. Network graph model for cohort 1 representing the individual level effects from the "influence" linear regression model for the positive affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

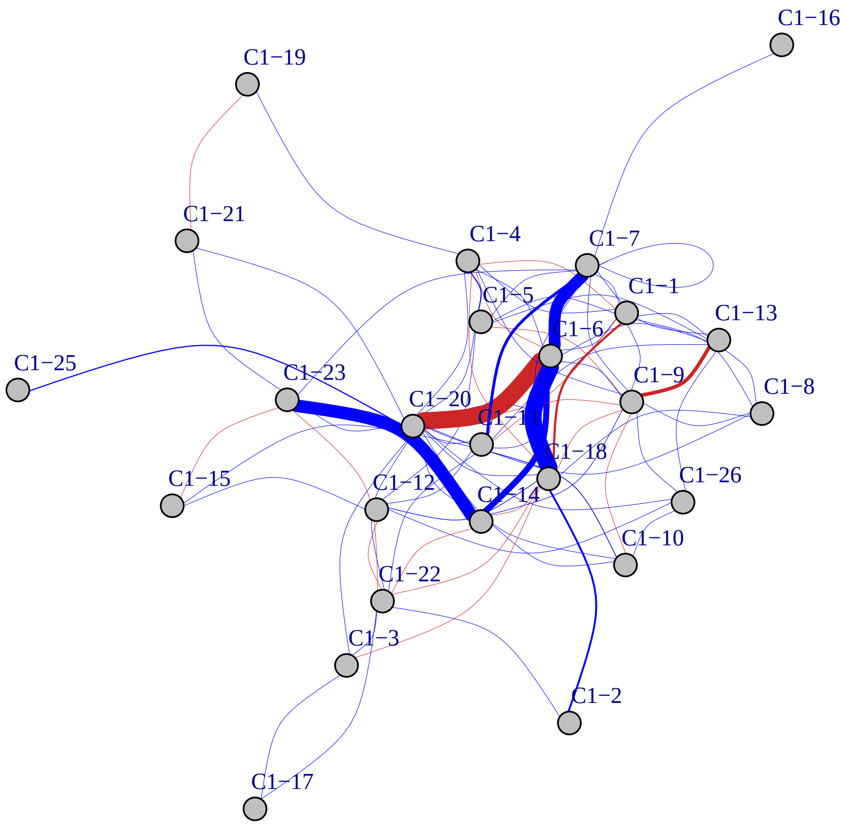

Figure 46. Network graph model for cohort 1 representing the individual level effects from the "influence" linear regression model for the negative affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

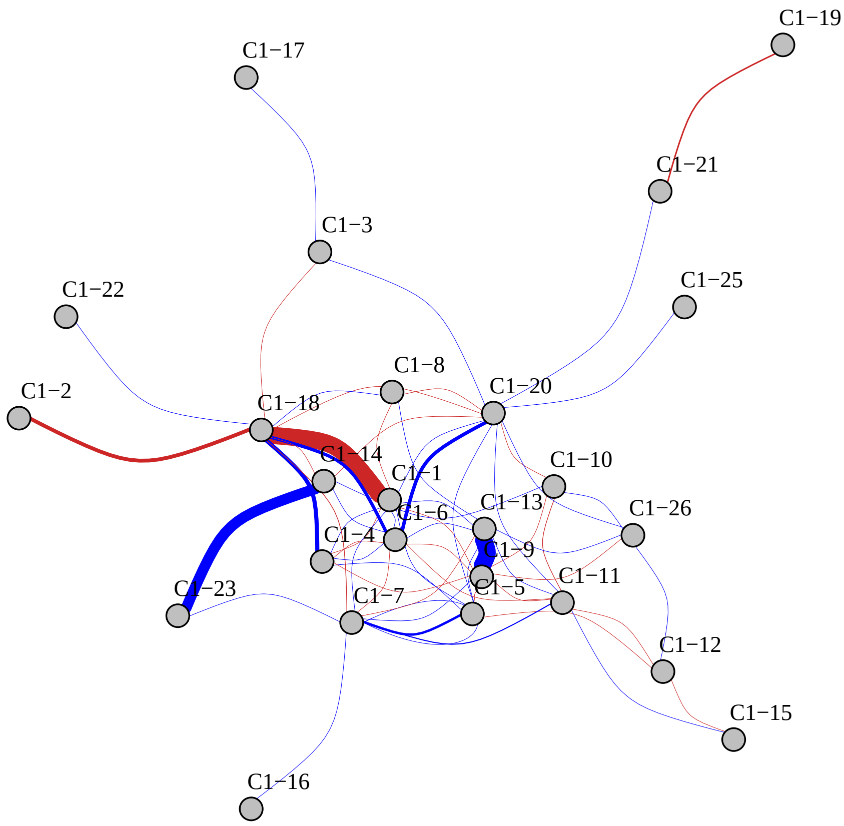

Figure 47. Network graph model for cohort 1 representing the individual level effects from the "influence" linear regression model for the alcohol survey response. Values that are in red are negative influence, values in blue are positive interactions.

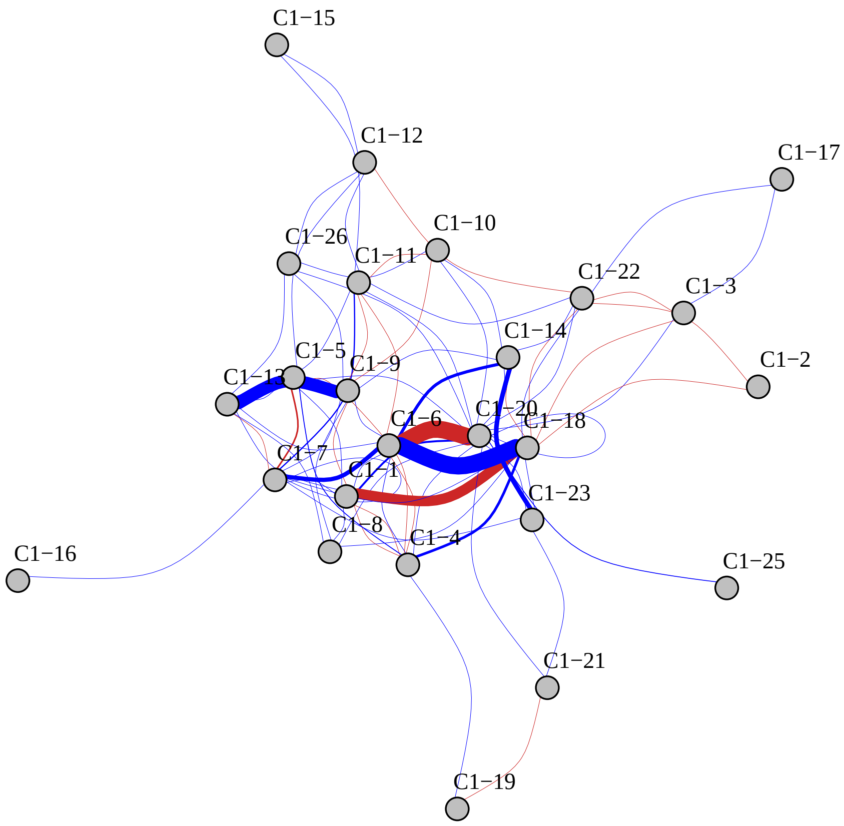

Figure 48. Network graph model for cohort 1 representing the individual level effects from the "influence" linear regression model for the sleep survey response. Values that are in red are negative influence, values in blue are positive interactions.

To be Findable:

F1. (Meta)data are assigned a globally unique and persistent identifier
F2. Data are described with rich metadata (defined by R1 below)
F3. Metadata clearly and explicitly include the identifier of the data it describes
F4. (Meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
  A2.1 The protocol is open, free, and universally implementable
  A2.2 The protocol allows for an authentication and authorization procedure, where necessary
A2. Metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (Meta)data use vocabularies that follow FAIR principles
I3. (Meta)data include qualified references to other (meta)data

To be Reusable:

R1. (Meta) data are richly described with a plurality of accurate and relevant attributes
  R1.1 (Meta)data are released with a clear and accessible data usage license
  R1.2 (Meta)data are associated with detailed provenance
  R1.3 (Meta)data meet domain-relevant community standards

Figure 49. FAIR Guiding Principles for scientific data management and stewardship

6 DATA NAVIGATION

6.1 Overview

The challenges of working with big data are often characterized in terms of 4 Vs: volume, variety, velocity, and veracity[8]. The MOSAIC performer data sets we were provided exemplify these characteristics. They are large (terabytes), consist of data drawn from a wide variety of sensors, and in some cases are high "velocity" in the sense that sensors sampled user activity at a very high rate. Veracity is also a potential issue, as our data analysis revealed patterns that suggest potential variability or inconsistency in how some sensors picked up human activity. As big data sets have become increasingly common, there is increasing consensus in the research community that better data management and governance will be crucial to making effective use of data. One standard in this area that is becoming increasingly relevant is the FAIR Guiding Principles for scientific data management and stewardship[14]. FAIR is an acronym that stands for four characteristics of data sets that are optimized for ease of use and reuse: Findable, Accessible, Interoperable, and Reusable. The complete set of guiding principles is provided in Figure 49. While it may not be practical for the MOSAIC project or its successors to adhere to all of these standards, attention to some of these principles may facilitate future collaborations around MOSAIC data sets. During the analysis, we encountered several issues relating to these principles:

- There were three main projects that captured and archived data from wearable sensors. Each had their own array of sensors and methods of collection. Each had their own schemas and structures in which data were captured.

- The data for the most part were not self-describing, with analysts having to rely on written documentation of varying quality and breadth to understand what data had been collected, where to find it, and what format to expect. (Partly because it was the best-documented and best-structured of the three collections, we chose to focus on the Notre Dame collection.)
- The data in the Notre Dame collection was typically not raw, but comprised one or more generations of derived data. While it was helpful to be able to work with pre-processed data, some tables (datasets) represented multiple versions of nearly the same derived datasets making it unclear which derived quantity and which version one should prefer for a given analysis.

6.2 MOSAIC data framework

Our data navigation efforts began with outlining a knowledge model or ontology that could be used to make MOSAIC data more findable for collaborators. Such a model needs to capture how data was collected and processed, and how it is intended to be understood – and further inform the analyst of its whereabouts and detailed structure. This model is based on several key abstractions, described below. The intention of this initial phase was to develop the model to a stage where sponsors and other participants would be able to understand how it could be incorporated into tools to help data analysts better relate to the MOSAIC data collections, and how one might go about developing those tools. The essential idea of our approach is to help analysts find the data they need by linking key concepts and relationships to the particular structures in the collection where associated data can be found. At the conclusion of this section, we provide an initial sketch of what a MOSAIC navigator tool might look like.

6.2.1 Abstractions for sensors and signals

These main concepts are ones familiar in signal processing relating to sensors and the signals they produce[2, 4, 3]. They are as follows:

- Sensor: Something that is a source of signals.
- Subject: The subject being observed by the sensor.
- Channel: The type of measurement provided by a sensor.
- Signal: A measurement, or series of measurements, recorded by a sensor on a given channel.
- Signature: A recognizable pattern embedded in a signal.

As an example, consider a measurement of heart rate supplied by a wearable device. The sensor is the wearable device. The subject would be the person wearing the sensor. Heart rate would be a channel on the sensor. The signal would be a measurement of heart rate or a series of such measurements, for example, heart rate over time. A signature would the portion of the signal that could be recognized as, say, an episode of accelerated heart rate, perhaps an indication of excitement or anxiety.

6.2.2 Abstractions for data structures

Given that signals embedded in MOSAIC datasets of interest, the signal processing abstractions suggested above can indicate to the analyst that a collection contains the desired signal. The next step is to find it, and to describe to the analyst, or to an application interfacing to the data collection, what the structure of the data is and how to read it. This means that our framework must be able to map the higher-level concepts that describe the signal to the lower-level concepts that describe data structures and schemas. For this we use some common concepts associated with relational database schemas. These concepts are generally applicable whether the data is actually stored in a relational database or not. They are:

collection relation property type domain description

SOURCE	TABLE/ CSV FILE	VARIABLE	DATA TYPE	FORMAT	MISSING VALUES	RANGE OF VALUES	DESCRIPTION
PA - Garmin HR	hr_series	snapshot_id	NUMERIC	N/A	N/A	NA	The ID of the participant
PA - Garmin HR	hr_series	day	DATE	yyyy-MM-dd	N/A	duration of study	Date for this series
PA - Garmin HR	hr_series	algorithm	SMALLINT	N/A	N/A	0 or 1	Algorithm to use for the hr_series calculation. 0 = Default, 5 minute window, 1 minute sliding interval. 1 = 5 minute window, 5 minute sliding interval.
PA - Garmin HR	hr_series	start_time	TIMESTAMP WITHOUT TIMEZONE	yyyy-MM-dd HH:mm:ss	N/A	within day of this entry	Start time for this HRV window
PA - Garmin HR	hr_series	stop_time	TIMESTAMP WITHOUT TIMEZONE	yyyy-MM-dd HH:mm:ss	N/A	within day of this entry	Stop time for this HRV window

Figure 50. An excerpt from the Notre Dame collection's data dictionary.

- Collection: A named collection of relations or tables.
- Relation: A set of related attributes or properties, i.e. a table.
- Property: An attribute or field in a table.
- Type: The type of allowed value for a field.
- Domain: The range of allowed values for a field.
- Description: A description of the field, meaning, or semantics.

Figure 50 illustrates these concepts using an excerpt from the Notre Dame data dictionary. Here we have taken the fields used in the Notre Dame schema and associated each of these concepts with a subset of them. The fields shaded blue were judged to be useful, but non-essential for our purposes.

6.2.3 MOSAIC-specific abstractions

There are concepts specific to the domain of the MOSAIC project that are not as abstract as the general signal analysis concepts presented above, but nonetheless important for narrowing down general concepts to specific ones like "sensor" to less specific ones like "beacon." By considering some of the documentation provided, we have arrived at the following sub-concepts to categorize datasets referenced in Notre Dame's data dictionary, and in their document of derived datasets. Figure 51 represents a start on a knowledge model/ontology for MOSAIC that could be used to navigate to specific datasets in the collection starting from very general concepts.

6.3 MOSAIC navigator application concepts

Our intention during this phase was to explore ideas for a navigator application, based on the semantic framework presented above. We envision the navigator as something like an ontology-enabled data browser that connects concepts like "performance monitor" and "subject heart rate" to properties like `snapshot_id`, `start_time`, and `stop_time` in table `hr_series` in the Notre Dame database.

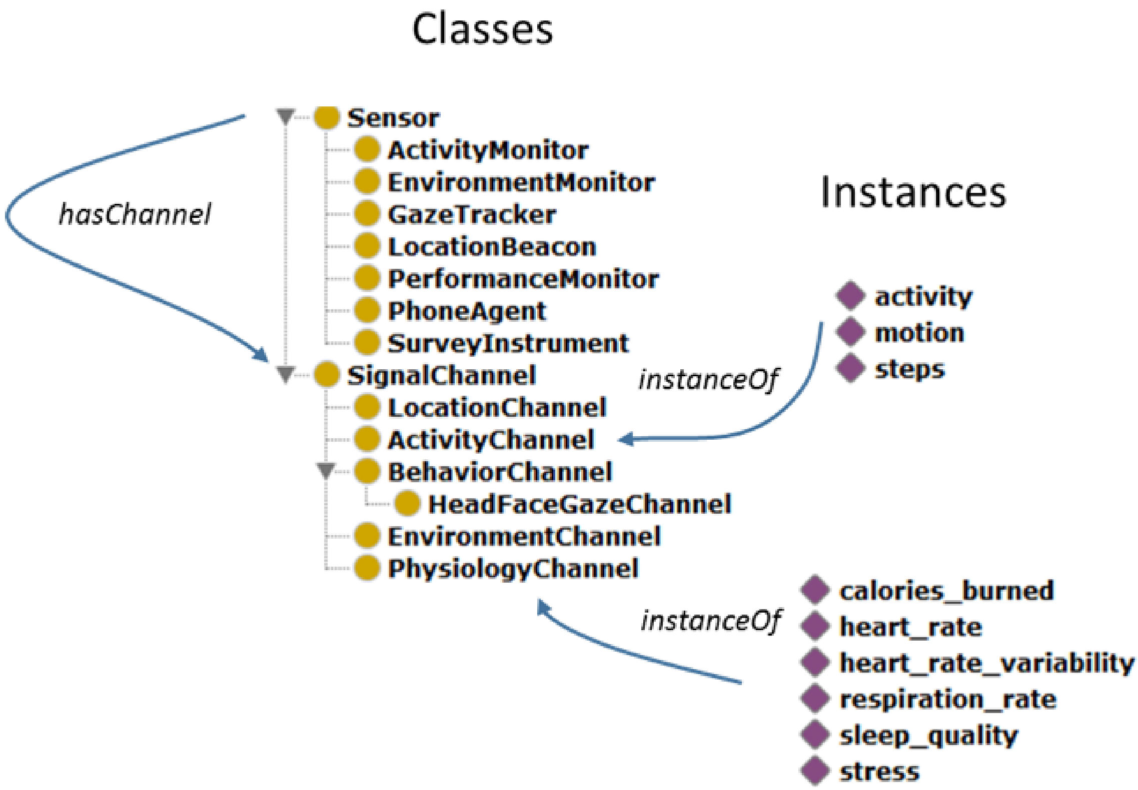

Figure 51. A partial knowledge model/ontology for abstractions relevant to MOSAIC data collections.

To explore the browser concept at the data structures level, we created a representation of the Notre Dame data dictionary in JSON format as a stub for a data storage layer that would provide information about data tables in the collection.

This representation involved several process steps that proved to be essential for capturing the dictionary in a clean, usable form for browsing. They included:

- Examining the data dictionary, which was provided in the form of an Excel spreadsheet.
- Querying the Notre Dame database, the form in which the actual tables were provided to us, to find the actual tables and fields of the data schema.
- Harmonizing the two representations to ensure that the descriptions in the data dictionary and the database fields and tables were consistent.
- Creating a JSON representation of the data dictionary as a hierarchy of tables and fields.

Figure 52 is an excerpt from the JSON data dictionary file. This file contains a tree of all the tables and fields in the collection along with descriptions extracted from the original dictionary. In the envisioned browser, the analyst might start with the concept hierarchy of sensors and signal channels to discover which tables and fields are relevant to the analysis. This might proceed using a hierarchy of classes and instances in a pull-down interface like the one shown in Figure 53. To browse the metadata in the JSON dictionary, we created a simple demonstration application in JavaScript. From there, the browser would take the analyst to the data dictionary hierarchy where they could query on the table name or some portion of it. At this point the dictionary browser may appear like the demonstration browser in Figure 54. We also explored the possibility of implementing a search tool alongside the dictionary browser, which would make

```
[{
    "name": "ND Data Dictionary",
    "level": "root",
    "id": "root",
    "children": [
        {
            "name": "pa_bbi",
            "level": "relation",
            "id": "pa_bbi",
            "children": [
                {
                    "name": "id",
                    "level": "property",
                    "id": "pa_bbi.id",
                    "children": [
                        {
                            "name": "type: bigint",
                            "level": "meta",
                            "id": "pa_bbi.id.bigint"
                        },
                        {
                            "name": "description: No description.",
                            "level": "meta",
                            "id": "pa_bbi.id.No description."
                        }
                    ]
                },
                {
                    "name": "local_time",
                    "level": "property",
                    "id": "pa_bbi.local_time",
                    "children": [
                        {
                            "name": "type: timestamp without time zone",
                            "level": "meta",
                            "id": "pa_bbi.local_time.timestamp without time zone"
                        },
                        {
                            "name": "description: Time in user local time. Used across many tables of the DB.",
                            "level": "meta",
                            "id": "pa_bbi.local_time.Time in user local time. Used across many tables of the DB."
                        }
                    ]
                },
```


Figure 52. A portion of the JSON file modeling the hierarchy of tables, fields, and description annotations for the Notre Dame MOSAIC collection.

it possible to search for a term like "heart," then highlight and expand all branches of the browser relevant to that search term. This would enable an analyst to find a term anywhere in the hierarchy and trace its connections up and down the hierarchy.

6.4 Summary

Following the MOSAIC data sprint, we engaged in two parallel efforts to address data navigation concerns:

1. An exploration of the concepts and relationships needed to understand what data is present in the collection, in terms of sensors, signal channels, and so on, as well as concepts that describe specific data structures like tables.

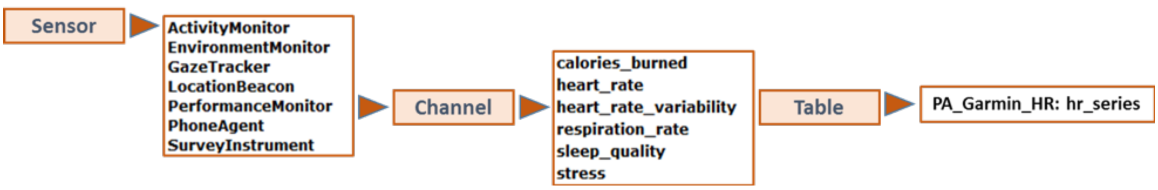

Figure 53. A concept for navigating from general concepts to tables within a data collection based on the abstract knowledge model.

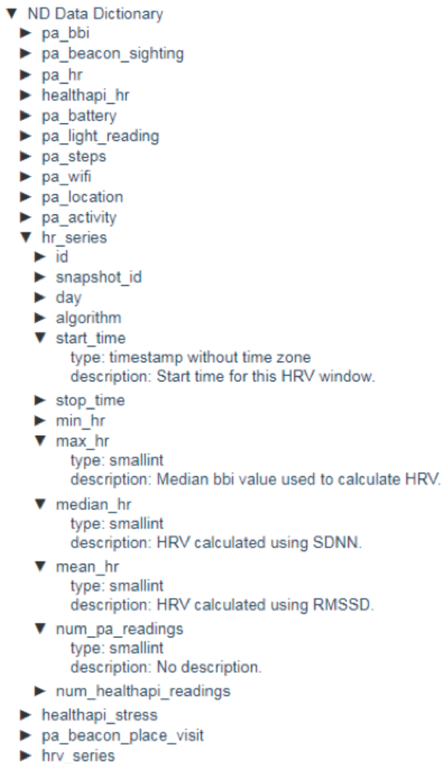

Figure 54. A screenshot from the JavaScript application demonstrating how an analyst would navigate through tables and fields in the data collection. This could be implemented alongside a search tool that would highlight search terms and automatically expand the relevant branches of the navigator.

2. An effort to write some of the code that might be used to implement the navigator as a web application.

Selecting the Notre Dame collection as our test case, we developed a partial knowledge model based on general concepts in signal analysis, as well as concepts related to domain-specific sensors and signals of in the IARPA MOSAIC project. This was done using documentation provided, including a data dictionary and a description of the main derived datasets, and querying the provided database for an inventory of tables and fields. An introductory presentation from the data sprint was also used as a source. To explore ideas for the browser, we created a representation of the harmonized data dictionary in JSON format to use as a stub for a persistent data tier in the actual application. We then created a hierarchical browser for the data dictionary in JavaScript. These explorations provide a conceptual basis for further development of the MOSAIC navigator concept that, according to our initial assessment, could be implemented as a web application. Although we used the Notre Dame data for this exploration because of its availability and documentation, the same process could be used to interface the other MOSAIC data collections to a navigator, with input from the other performers. The general abstractions provided by a knowledge model such as the one described here would make it possible to create a common interface to the current data collections, as well as future collections produced by similar efforts. Our experience suggests that such a tool is possible with modest effort, and that it could be of substantial benefit to analysts interested in exploring MOSAIC data collections.

7 CONCLUSIONS AND FUTURE WORK

7.1 Limitations of the analysis

As discussed in the introduction and elsewhere in this report, our limited access to organizational ground truth makes it difficult to assess the significance or validity of our results, and limits our ability to engage in the kind of iterative refinement of research directions and analysis methods that might otherwise characterize this type of analysis. More concretely, as a result of our analyses efforts, we note two key anomalies in the Notre Dame beacon data that may limit the accuracy of our analysis. First, there is significant asymmetry in interaction durations reported by beacons that were in contact with one another. That is, beacon A did not always report being in proximity to beacon B at the same time as B reported being in proximity to A. This is somewhat surprising, since mutual proximity is a transitive phenomenon. This suggests there may be unexplored sensor calibration issues or problems with inconsistent beacon usage. Second, in the Notre Dame processed interaction sessions data we used for most of our analyses, we find what appear to be unrealistically low total interaction times for participants, with the vast majority of participants reporting less than an hour of interaction with all other participants combined over the course of nearly a year. We chose to use the processed interaction sessions as our baseline because we believe the Notre Dame team may have had good reasons for throwing out some of the interaction data based on their knowledge of the data collection process. However, not knowing the specific assumptions behind their analysis leaves open the possibility that we are interpreting this data incorrectly, or basing our analysis on a biased sample of the data.

7.2 Data collection and management recommendations

Our experience analyzing the Notre Dame beacon data suggests several recommendations for future collection efforts aimed at providing data for social/organizational analysis. Some of these recommendations reflect the fact that the original data collection process was optimized for maximizing the number of participants and for analysis of individual behavior; data collection requirements are somewhat different for sociological analyses. Other recommendations aim at better characterizing the sensors, particularly in the context of daily use by participants. Finally, we note that more attention to data management and documentation would make these data sets much more usable, both for the performer teams and those they may share data with. Our recommendations for addressing these issues are summarized briefly below.

Sampling concerns

- Use more structured recruiting and sampling methods if possible
- Carefully document recruitment strategies
- Focus on complete coverage of organizational units rather than broad coverage of the organization

Additional organizational ground truth

- Organizational roles
- Organizational unit membership
- Events (project start/end, major deliverable dates)

Sensor reliability issues

- Experiment with the effects of different sensor configurations before deployment
- Study how participants actually use sensors in the field

- Validate sensor results against other observations in small studies

Enabling data sharing

- Follow standards for data management (formal data management plans, etc.)
- Emphasize documentation (data dictionaries, metadata, etc.)
- Allow time for pre-processing of data before sharing with others

7.3 Testable hypotheses

Our analyses have generated several concrete hypotheses that could in theory be tested by returning to the participants or their organizations. We present these here as generalized proposals for confirming our findings, although they can each be transformed into one or more specific hypotheses. Because we do not have ground truth regarding social groups or influence, we cannot yet validate either of our analyses, but investigation of these proposals could provide valuable information for assessing and refining our methods. We divide the proposals into three categories: organizational characteristics, group formation and dynamics, and influence.

Organizational characteristics

- Cohorts 2 and 4 appear to be very closely connected or possibly share one or more physical locations.
- Cohorts 2 and 3 appear to be connected, but to a much lesser extent than cohorts 2 and 4. This suggests these organizations have some interaction but may not be physically co-located
- Cohort 4 appears to be much less connected internally than the other cohorts (with the exception of cohort 5, which we already know is spread across multiple organizations/locations). This suggests that cohort 4 might represent a sparser sample of its organization than the other cohorts.

Group formation and dynamics

- In general, any of the groups we identify as sizable and persistent could be checked against organizational ground truth to see if they correspond to management units, projects, or shared work locations.
- Some change appears to occur around week 20 in cohort 1. This could reflect changes in data collection or participant compliance, or some shift in the work environment, such as a new project or a re-organization.
- In cohort 2, group 0, 1, and 6 form around week 20 and persist. Again, this could be a collection or compliance issue, or an actual workplace or social change.
- In cohort 3, groups 9 and 21 show significant change over the course of the study, in size and gender distribution. It could be confirmed whether something was changing in the organization socially or professionally for these specific people.
- Although the data are small, stress levels over time per group in cohort 2 might imply that smaller groups experience more volatile stress levels. It may be difficult, but follow-up interviews with cohort member may illuminate this.

Influence

- Any of the strong influence relationships identified between individuals could possibly be tested against survey questions asking about closeness of interpersonal connections.

- We could characterize work roles of highly influential individuals in more detail to see if there is any consistent association between role and influence. It may be possible to do this using existing Notre Dame data.

8 ACKNOWLEDGEMENTS

We thank Alexis Jeannotte of IARPA for supporting this research, Matt Heavner for LANL program support, and John Ambrosiano and John Whitton for assisting with data management. For data sprint planning and facilitation, we thank John Schoonover and Molly Cernicek, and Dan Padgett and Vel Preston of Air Force Cyberworx. Juston Moore led initial MOSAIC efforts at LANL, assisted with many aspects of the data sprint, and set up the computing interface for the sprint and later analysis work. James Wernicke and Will Rosenberger configured and maintained computing servers for the project. Many LANL MOSAIC Data Sprint participants also contributed ideas and methods that informed this analysis. From LANL, participants included Ashlynn Daughton, Diane Oyen, Jason Gans, Vlad Henzl, Leticia Cuellar, Nathan Lemons, Lori Dauelsberg, and Trinity Overmyer. External sprint participants included Alex Danvers from the University of Arizona, on behalf of the Lockheed-Martin team; Gina Notaro of Lockheed Martin; and Ann Gibbon of the University of Auckland and Matri Design. Finally, we thank the IARPA MOSAIC Performers (Lockheed Martin, the University of Notre Dame, and the University of Southern California), particularly Aaron Striegel and Gonzalo Martinez of the Notre Dame team, for briefings, advice, and access to their data.

We gratefully acknowledge funding support from the Multimodal Objective Sensing to Assess Individuals with Context (MOSAIC) program of the Intelligence Advanced Research Projects Agency, Office of the Director of National Intelligence. The views and conclusions contained herein are those of the authors, and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, LANL, DOE, or the US Government.

REFERENCES

[1]A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15. ACM, 2008.

[2]P. Barnaghi, S. Meissner, M. Presser, and K. Moessner. Sense and sens'ability: Semantic data modelling for sensor networks. In *Proceedings of the ICT-Mobile Summit*, 2016.

[3]M. Compton et al. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.

[4]M. Compton, C. Henson, L. Lefort, H. Neuhaus, and A. Sheth. A survey of the semantic specification of sensors. In *Proceedings of the 2nd International Conference on Semantic Sensor Networks*, pages 17–32, 2009.

[5]P. Dodds and D. J. Watts. Threshold models of social influence. In *The Oxford Handbook of Analytical Sociology*. 2009.

[6]M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[7]M. Granovetter and R. Soong. Threshold models of diffusion and collective behavior. *Journal of Mathematical sociology*, 9(3):165–179, 1983.

[8]R. Kitchin and G. McArdle. What makes big data, big data? exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1):1–10, 2016.

[9]S. M. Mattingly, J. M. Gregg, P. Audia, A. E. Bayraktaroglu, A. T. Campbell, N. V. Chawla, V. Das Swain, M. De Choudhury, S. K. D'Mello, A. K. Dey, G. Gao, K. Jagannath, K. Jiang, S. Lin, Q. Liu, G. Mark, G. J. Martinez, K. Masaba, S. Mirjafari, E. Moskal, R. Mulukutla, K. Nies, M. D. Reddy, P. Robles-Granda, K. Saha, A. Sirigiri, and A. Striegel. The Tesserae project: Large-scale, longitudinal, in situ, multimodal sensing of information workers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, page CS11. ACM, 2019.

[10]P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[11]A. Páez, D. M. Scott, and E. Volz. Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality. *Social Networks*, 30(4):309–317, 2008.

[12]T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 1978.

[13]J. Sun and J. Tang. A survey of models and algorithms for social influence analysis. In *Social network data analytics*, pages 177–214. Springer, 2011.

[14]M. D. Wilkinson et al. The FAIR guiding principles for scientific data management and stewardship. *Nature Scientific Data*, 3(160018), 2016.

APPENDIX A DATA USED FOR DYNAMIC GROUP ANALYSIS PARAMETER SELECTION

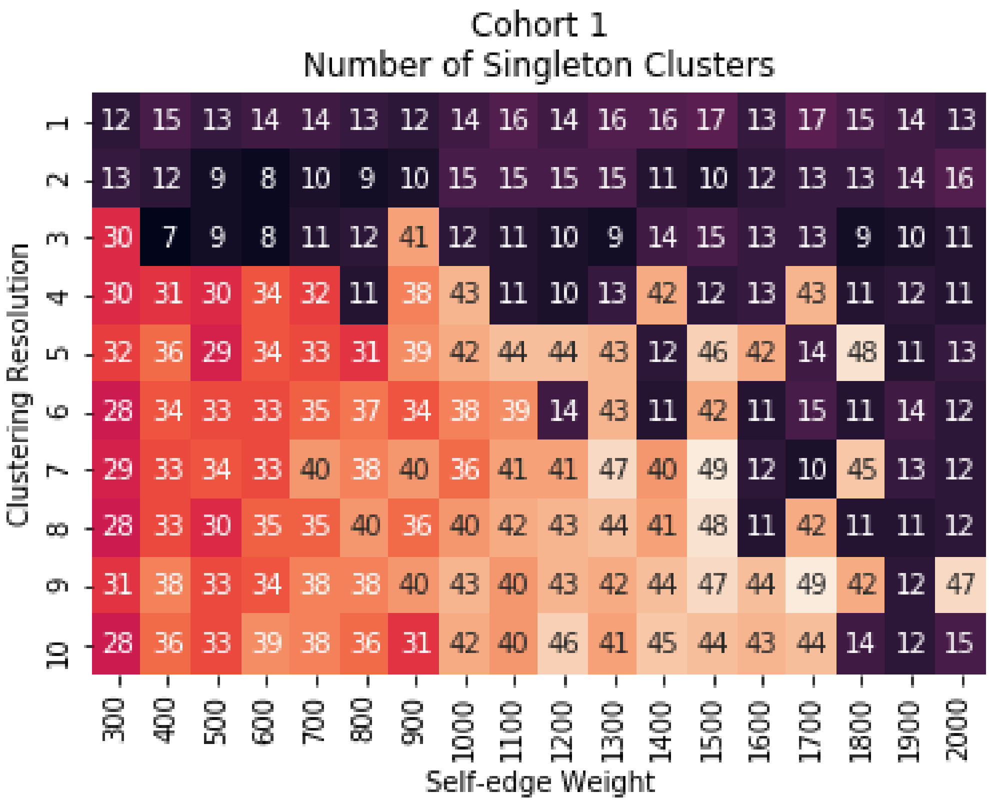

Figure 55. Number of singleton clusters in cohort 1.

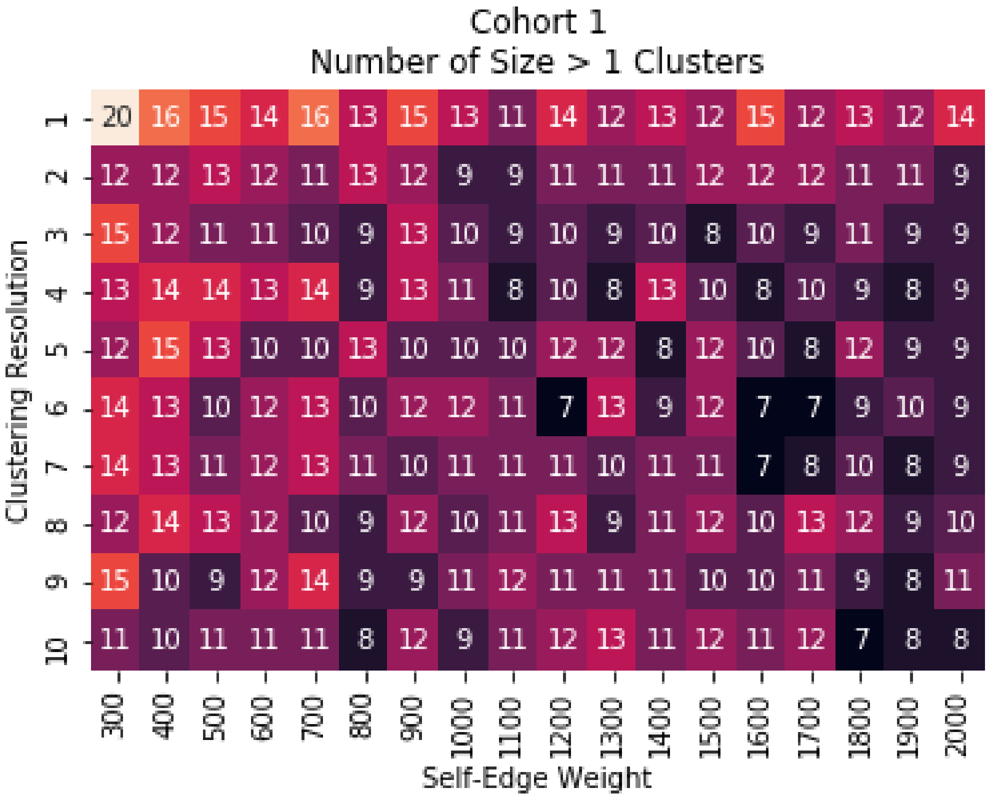

Figure 56. Number of size > 1 clusters in cohort 1

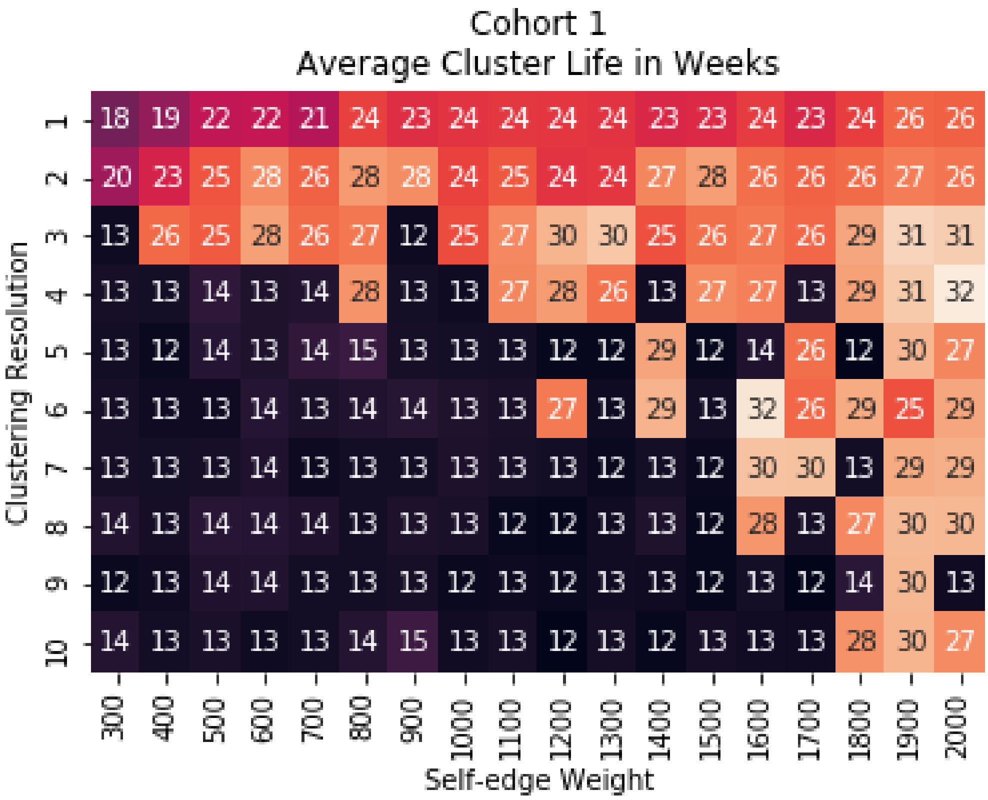

Figure 57. Average cluster life in weeks in cohort 1.

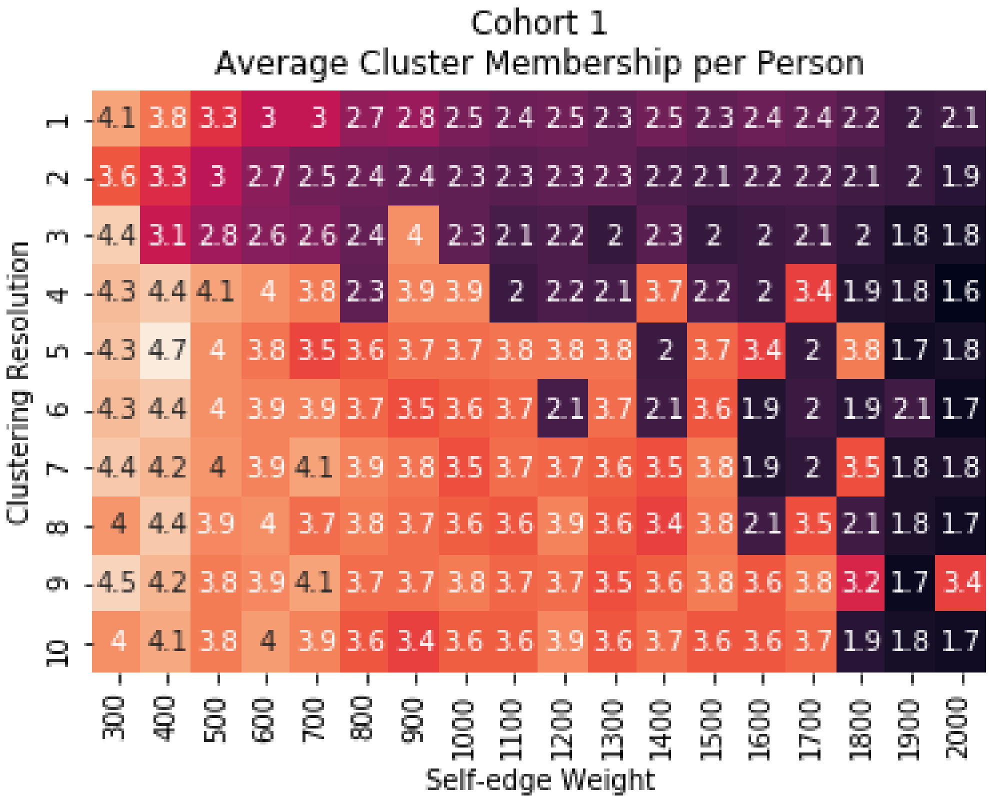

Figure 58. Average number of clusters each participant belongs to over the course of the study in cohort 1.

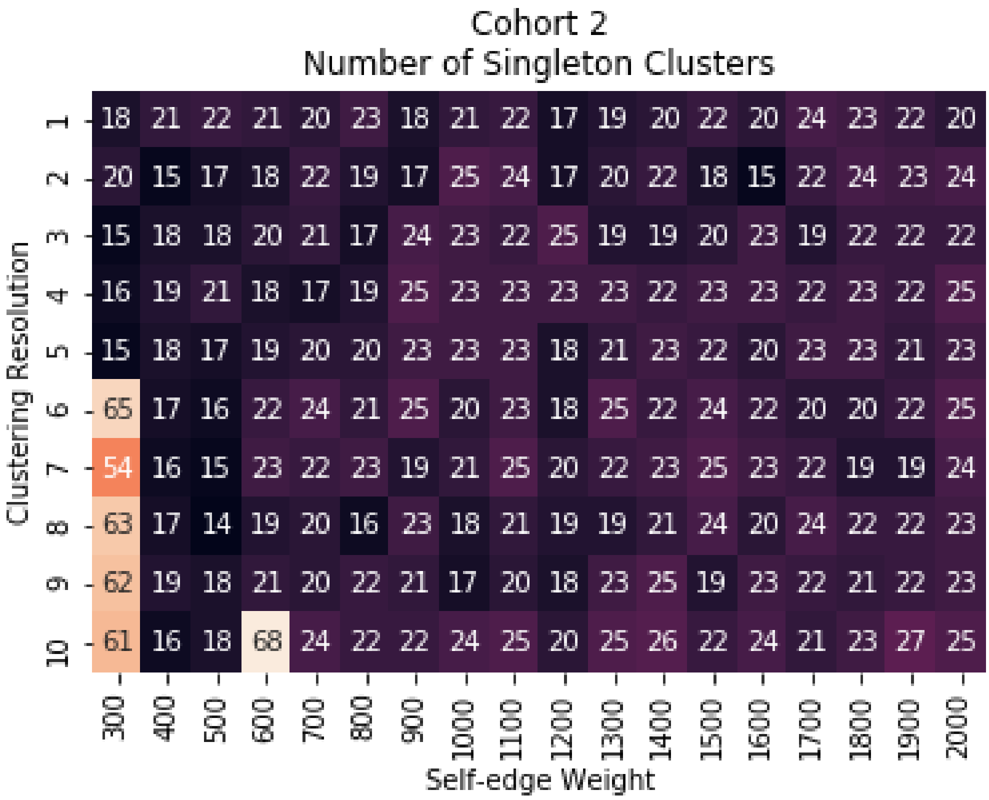

Figure 59. Number of singleton clusters in cohort 2.

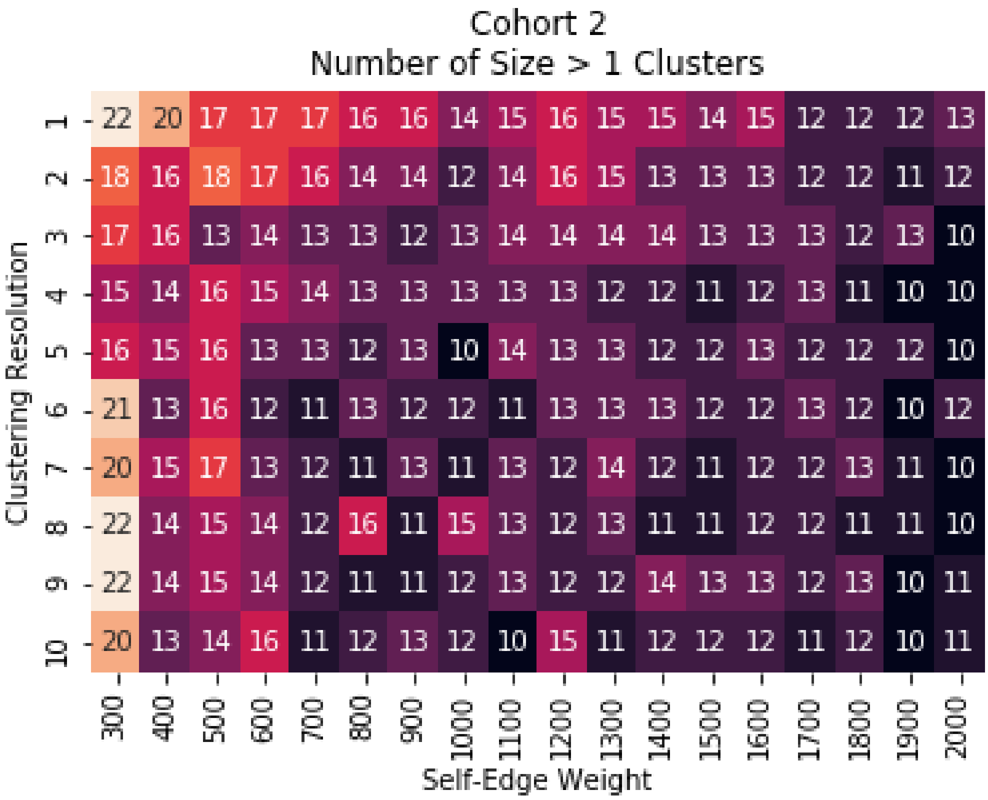

Figure 60. Number of size $>$ 1 clusters in cohort 2.

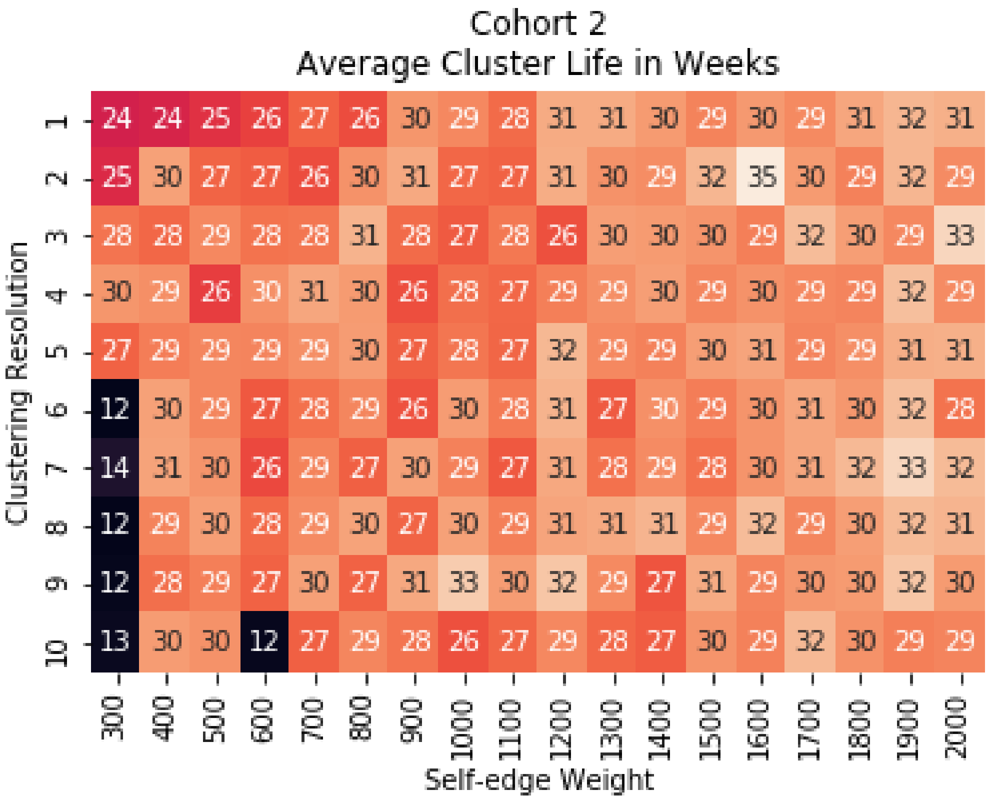

Figure 61. Average cluster life in weeks in cohort 2.

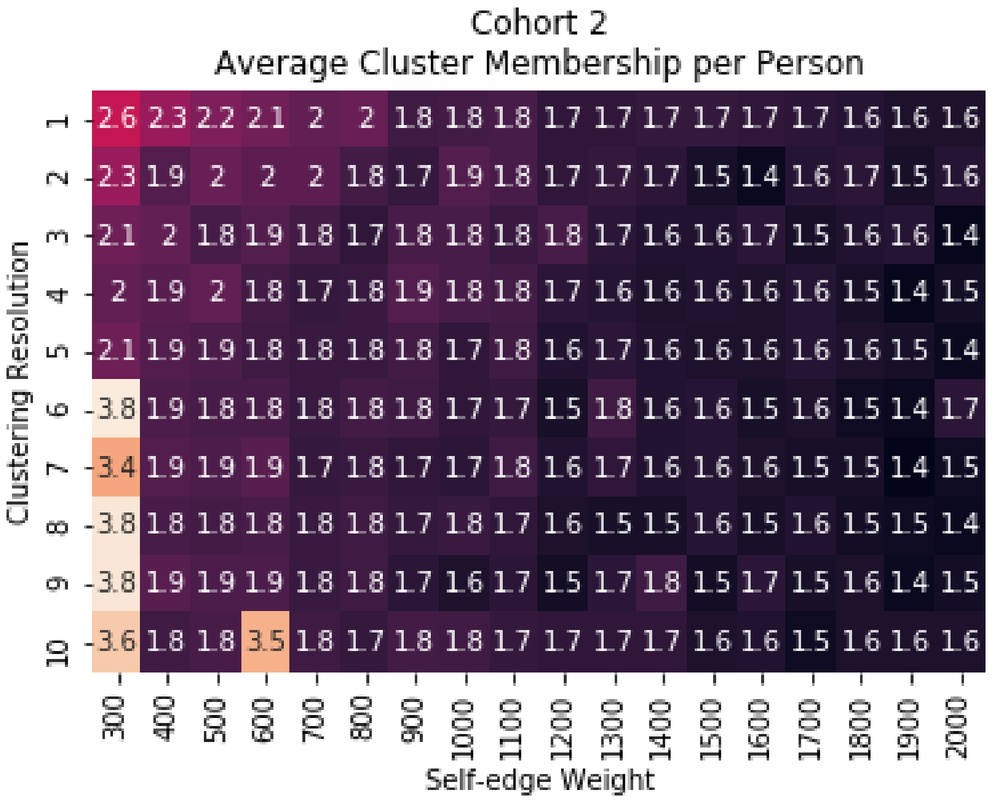

Figure 62. Average number of clusters each participant belongs to over the course of the study in cohort 2.

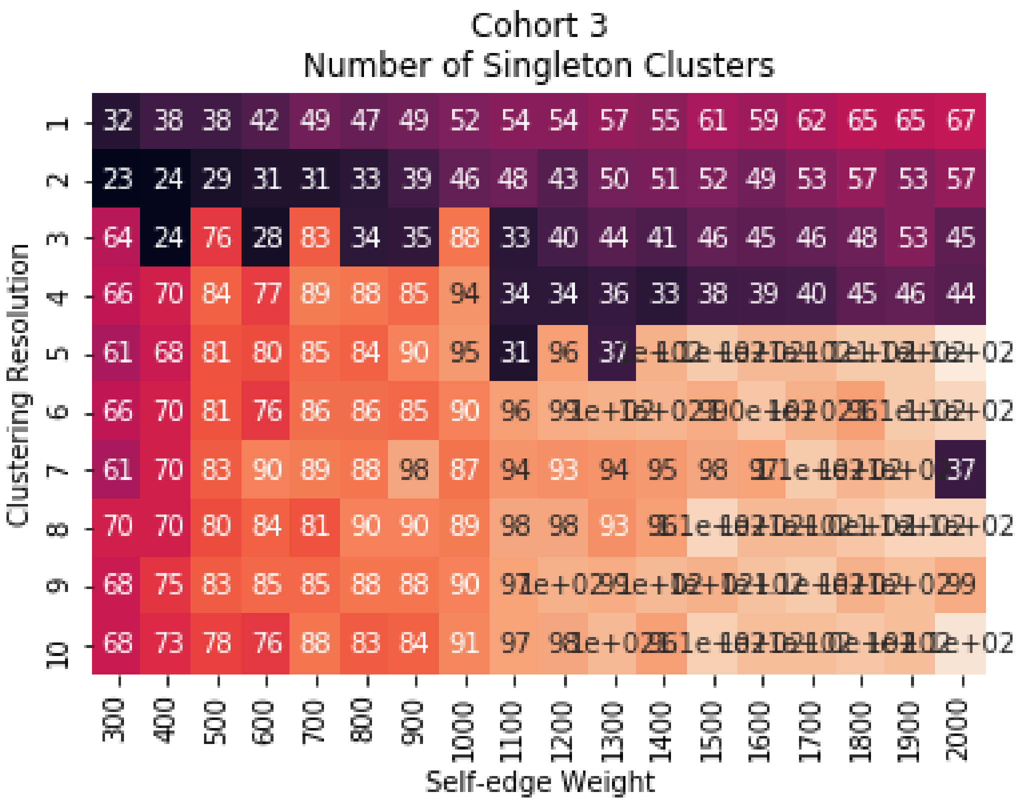

Figure 63. Number of singleton clusters in cohort 3.

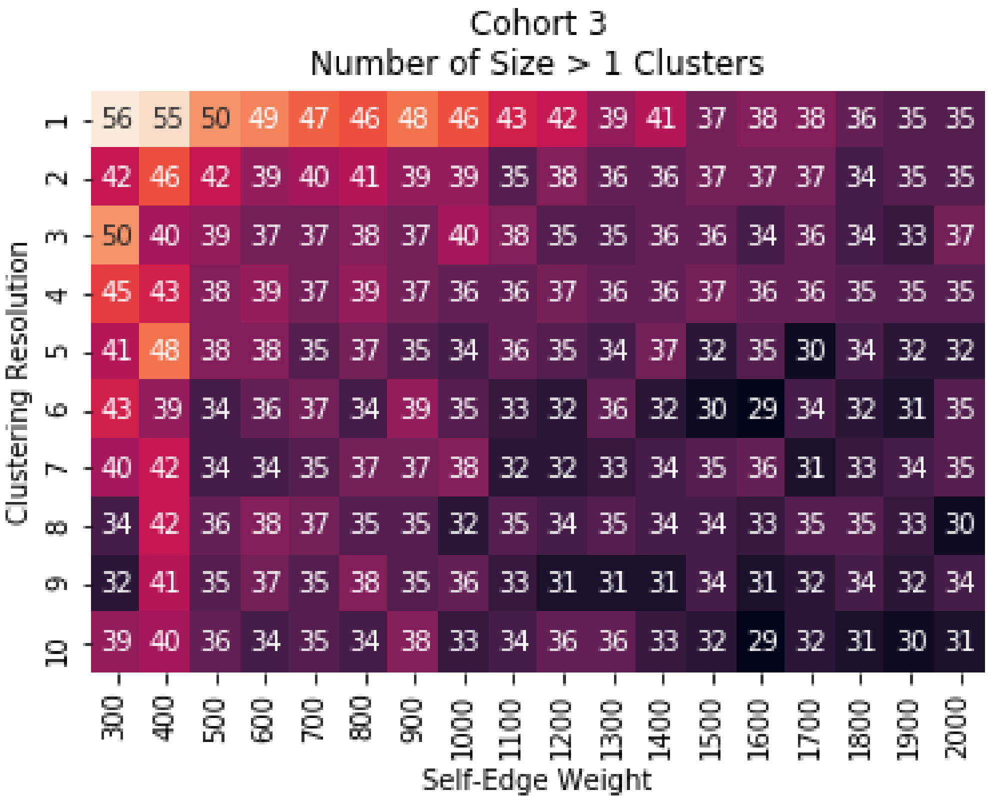

Figure 64. Number of size > 1 clusters in cohort 3.

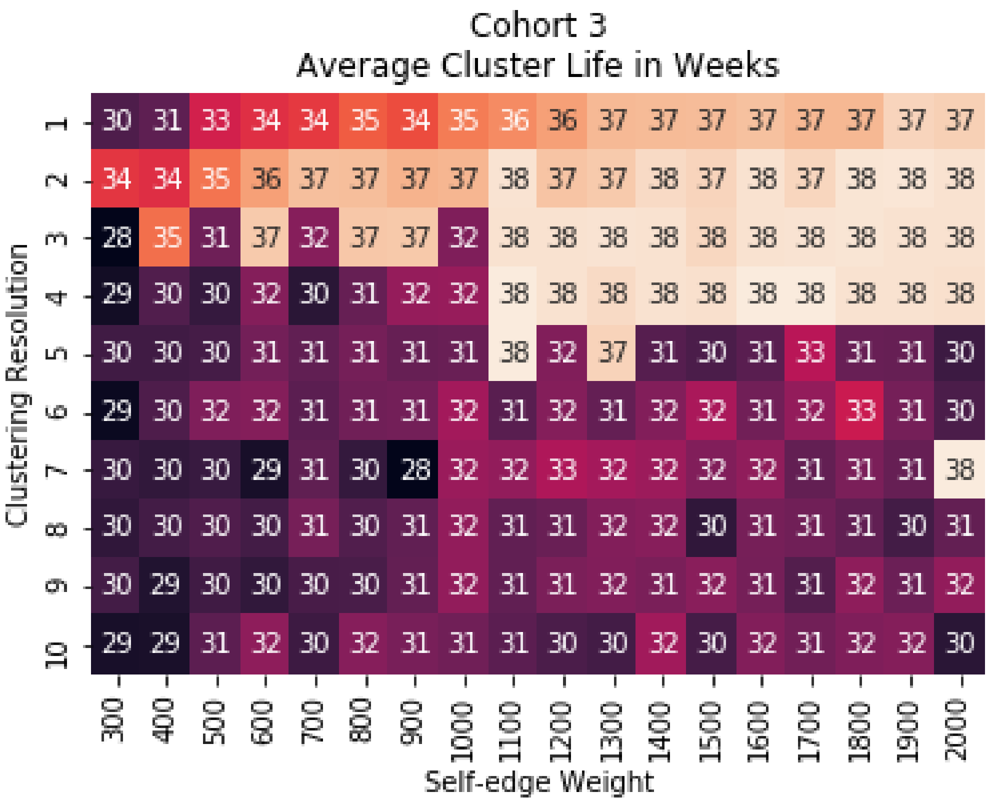

Figure 65. Average cluster life in weeks in cohort 3.

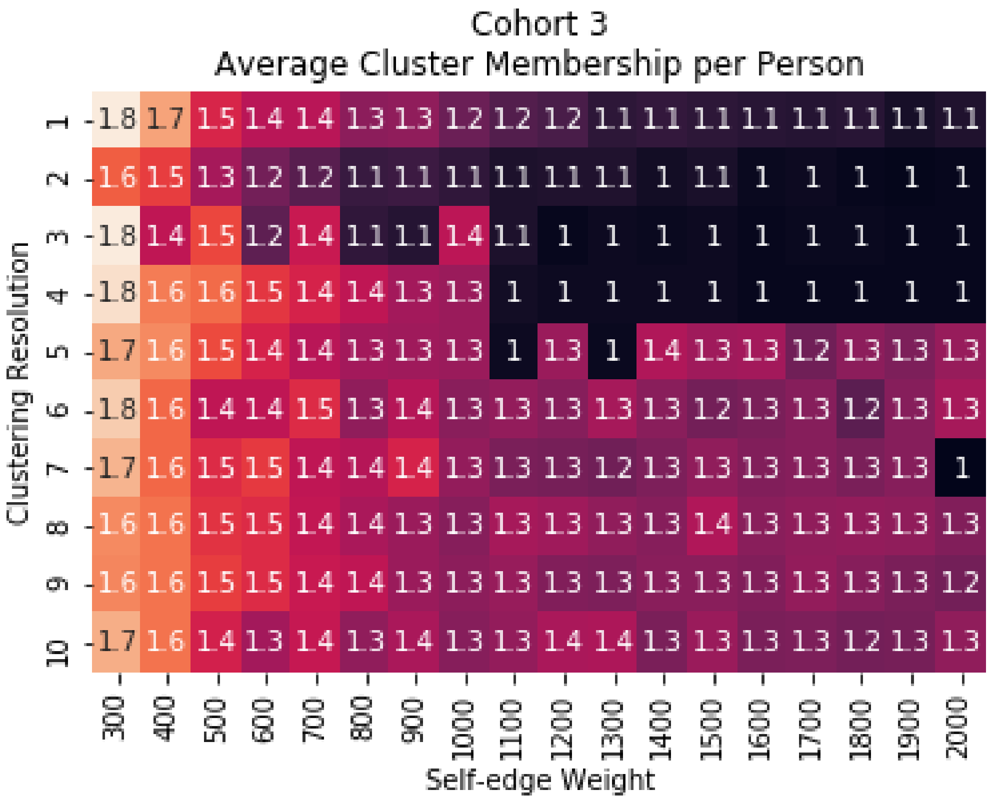

Figure 66. Average number of clusters each participant belongs to over the course of the study in cohort 3.

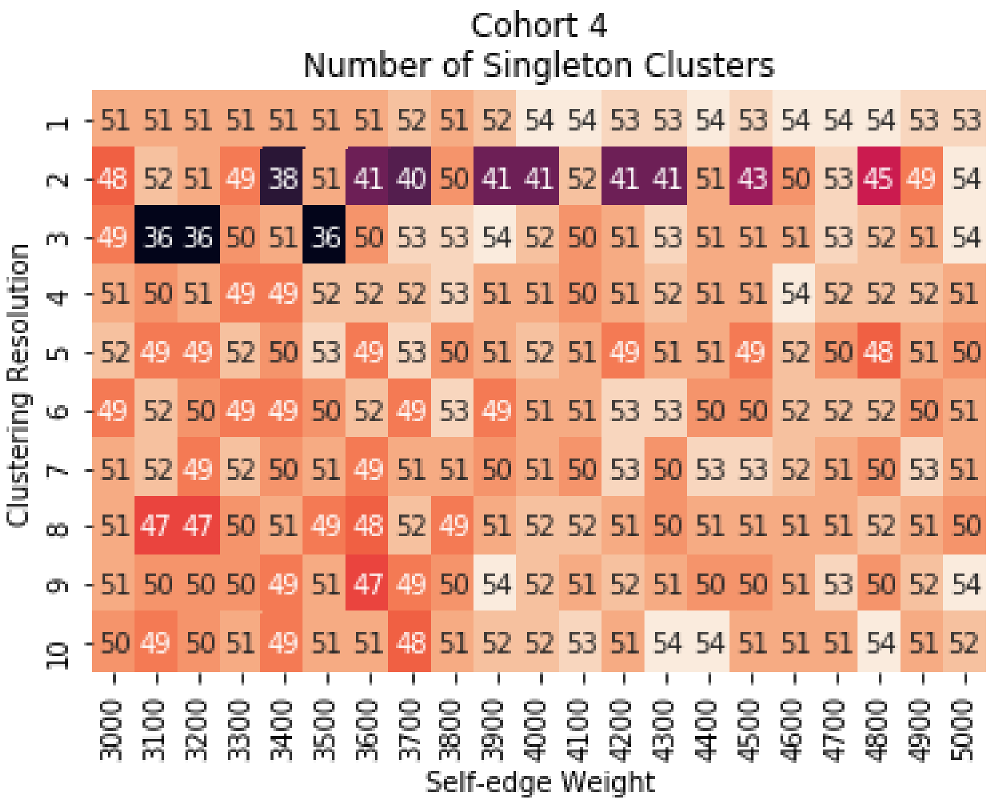

Figure 67. Number of singleton clusters in cohort 4.

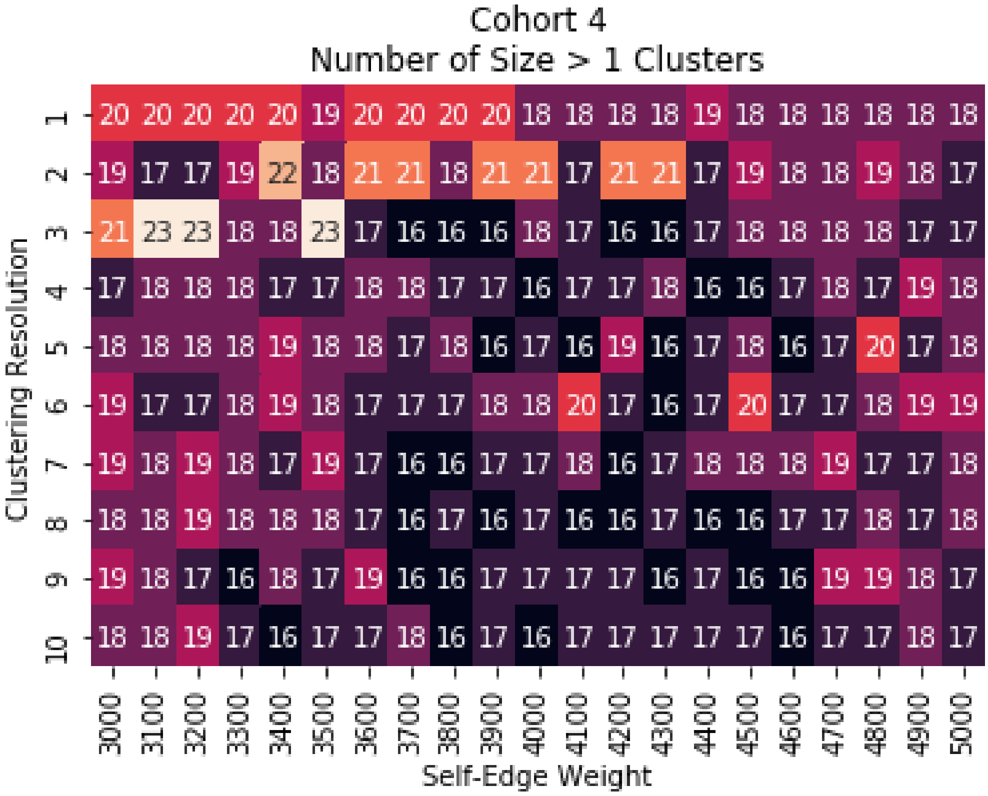

Figure 68. Number of size $>$ 1 clusters in cohort 4.

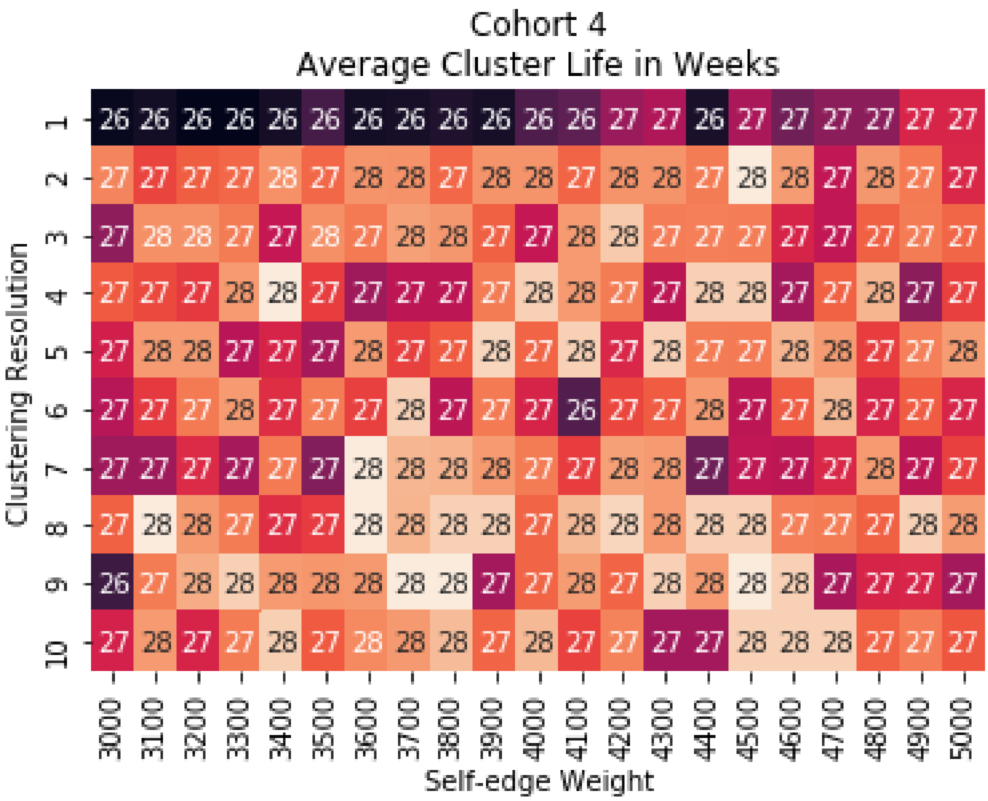

Figure 69. Average cluster life in weeks in cohort 4.

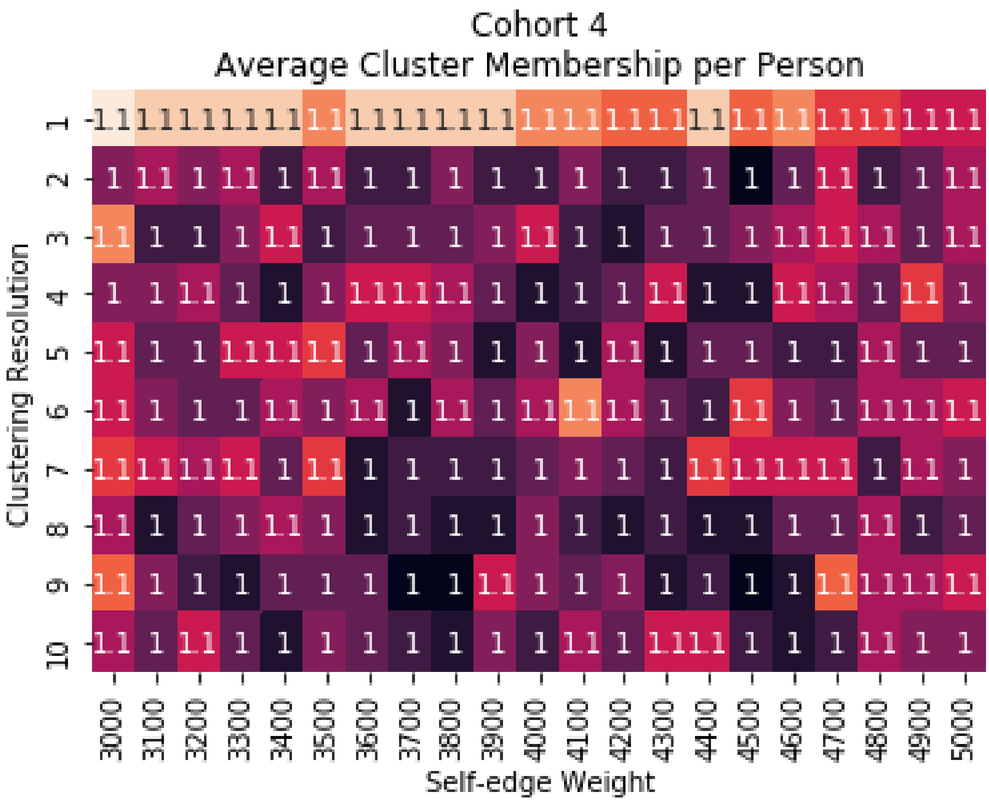

Figure 70. Average number of clusters each participant belongs to over the course of the study in cohort 4.

APPENDIX B ADDITIONAL INFLUENCE ANALYSIS FIGURES

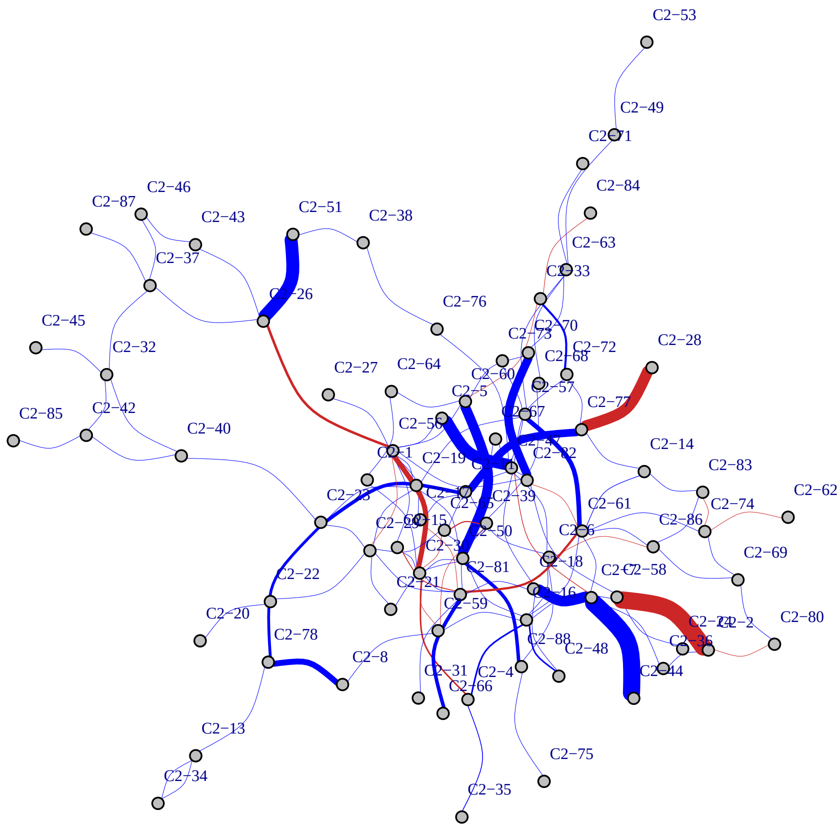

Figure 71. Network graph model for cohort 2 representing the individual level effects from the “influence” linear regression model for the positive affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

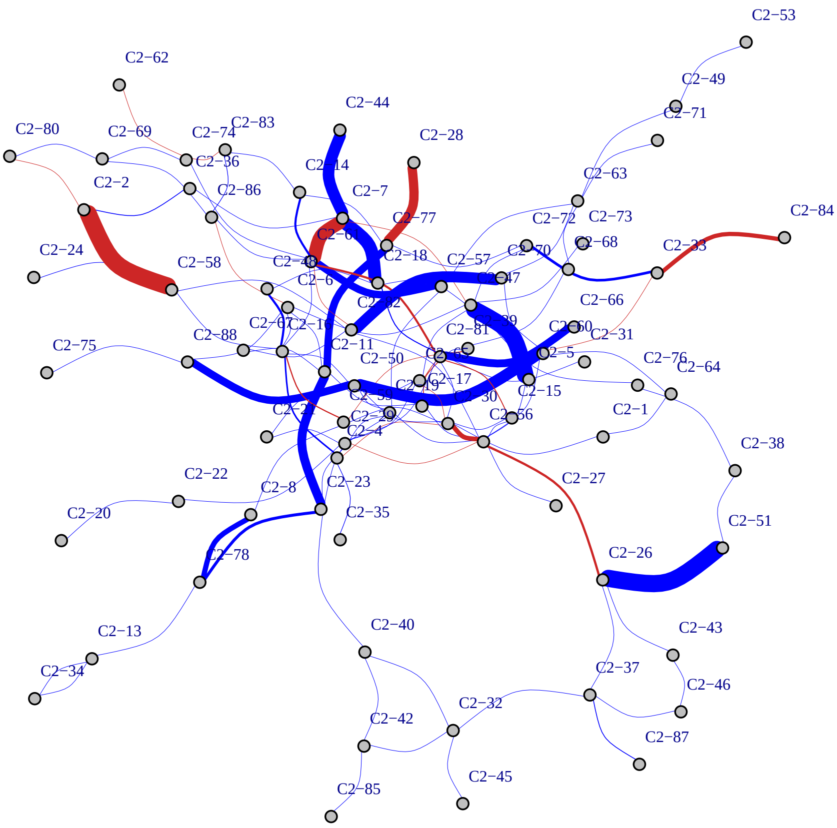

Figure 72. Network graph model for cohort 2 representing the individual level effects from the "influence" linear regression model for the negative affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

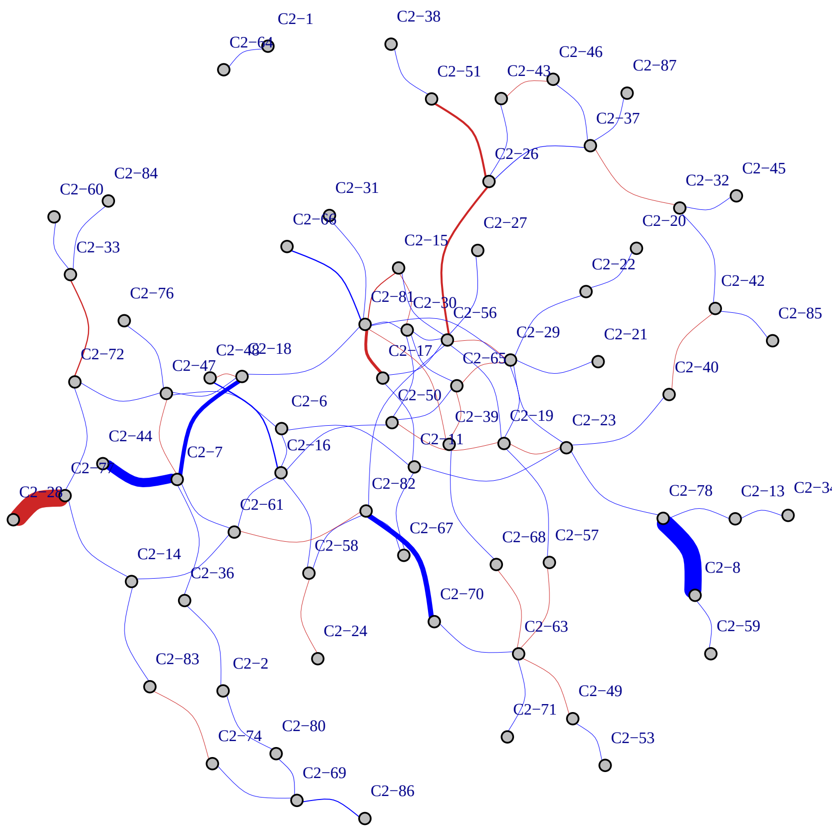

Figure 73. Network graph model for cohort 2 representing the individual level effects from the "influence" linear regression model for the alcohol survey response. Values that are in red are negative influence, values in blue are positive interactions.

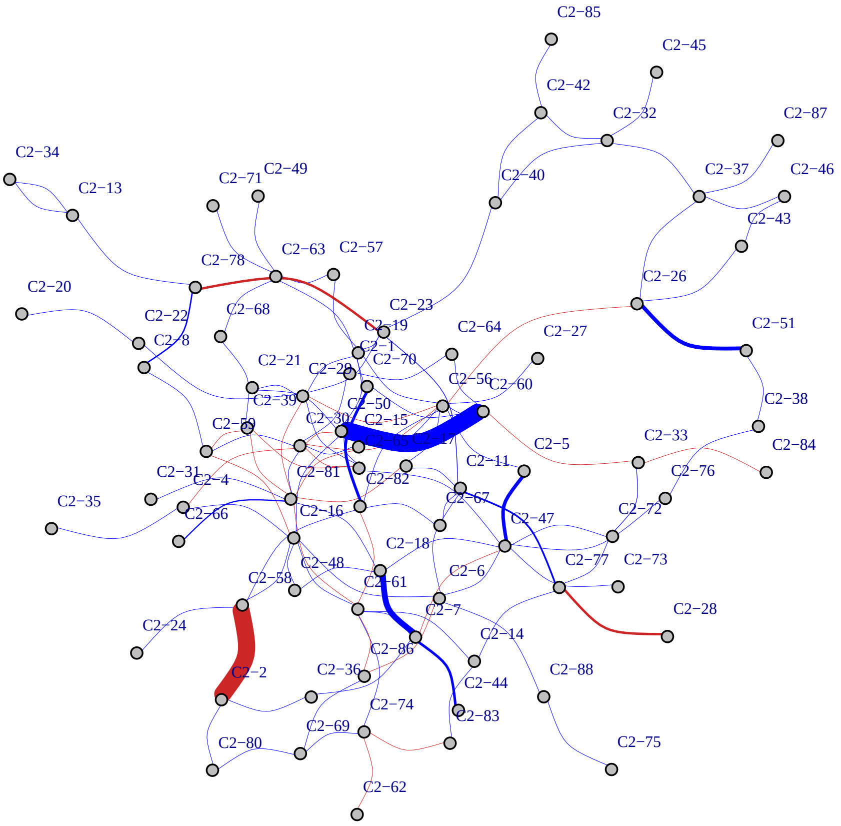

Figure 74. Network graph model for cohort 2 representing the individual level effects from the "influence" linear regression model for the sleep survey response. Values that are in red are negative influence, values in blue are positive interactions.

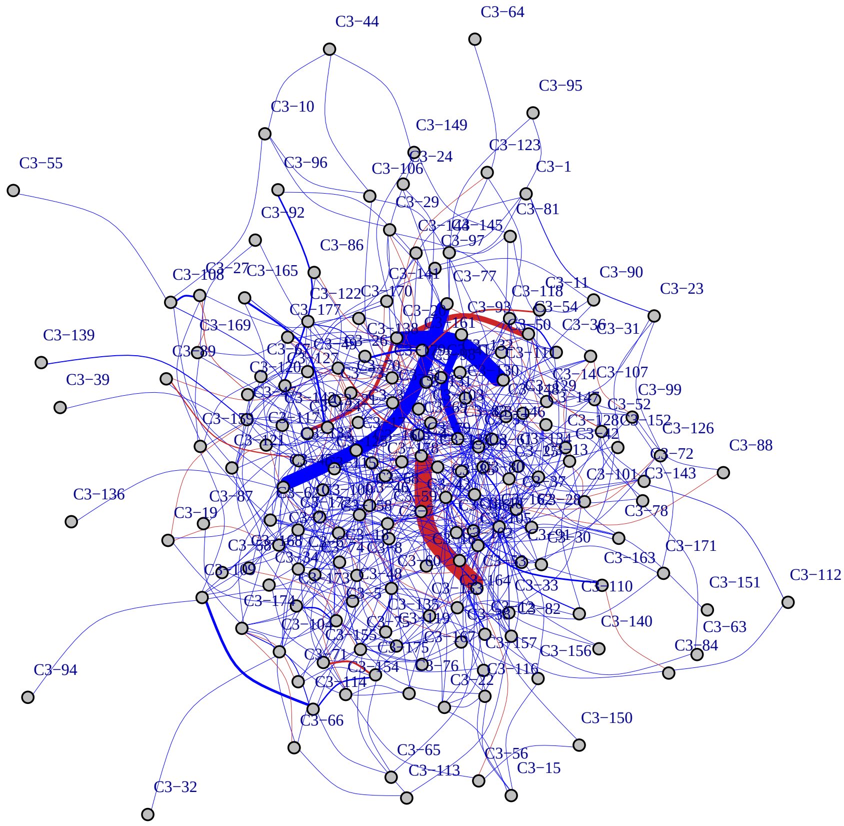

Figure 75. Network graph model for cohort 3 representing the individual level effects from the "influence" linear regression model for the positive affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

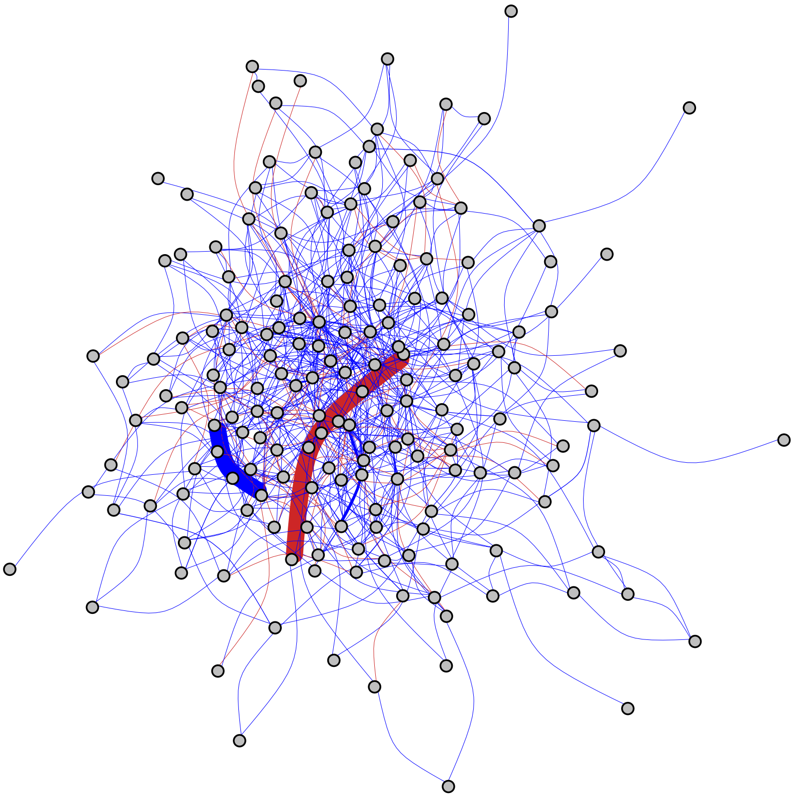

Figure 76. Network graph model for cohort 3 representing the individual level effects from the “influence” linear regression model for the negative affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

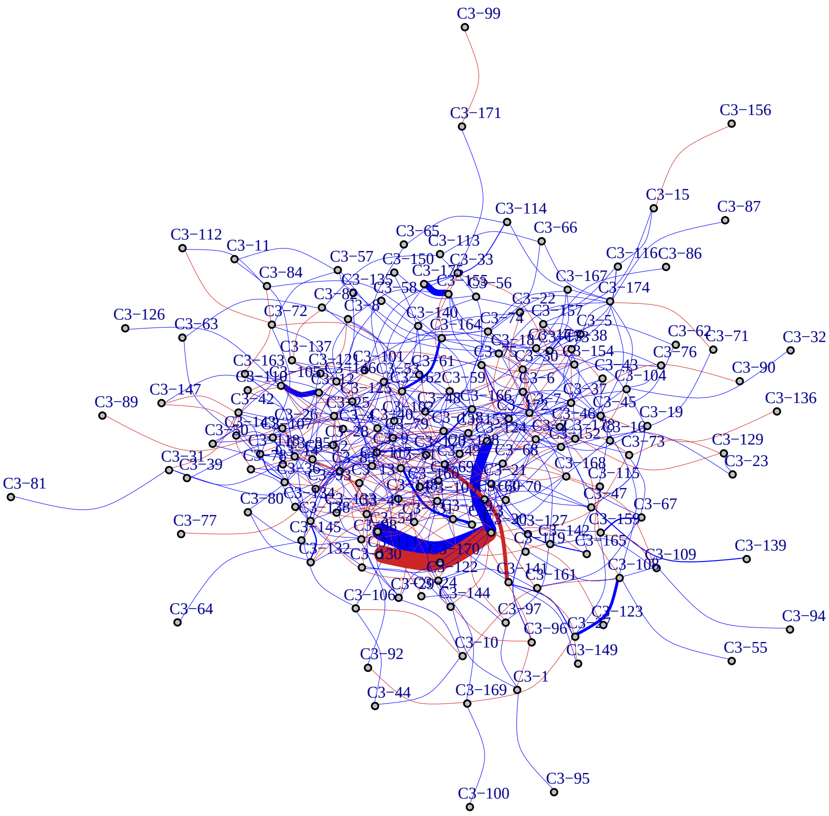

Figure 77. Network graph model for cohort 3 representing the individual level effects from the "influence" linear regression model for the alcohol survey response. Values that are in red are negative influence, values in blue are positive interactions.

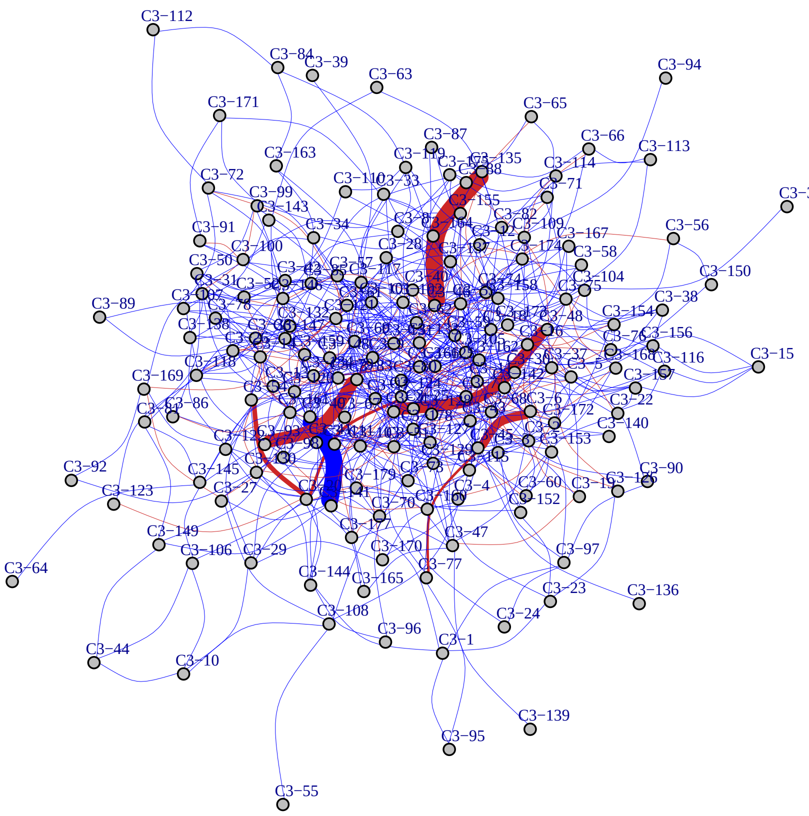

Figure 78. Network graph model for cohort 3 representing the individual level effects from the “influence” linear regression model for the sleep survey response. Values that are in red are negative influence, values in blue are positive interactions.

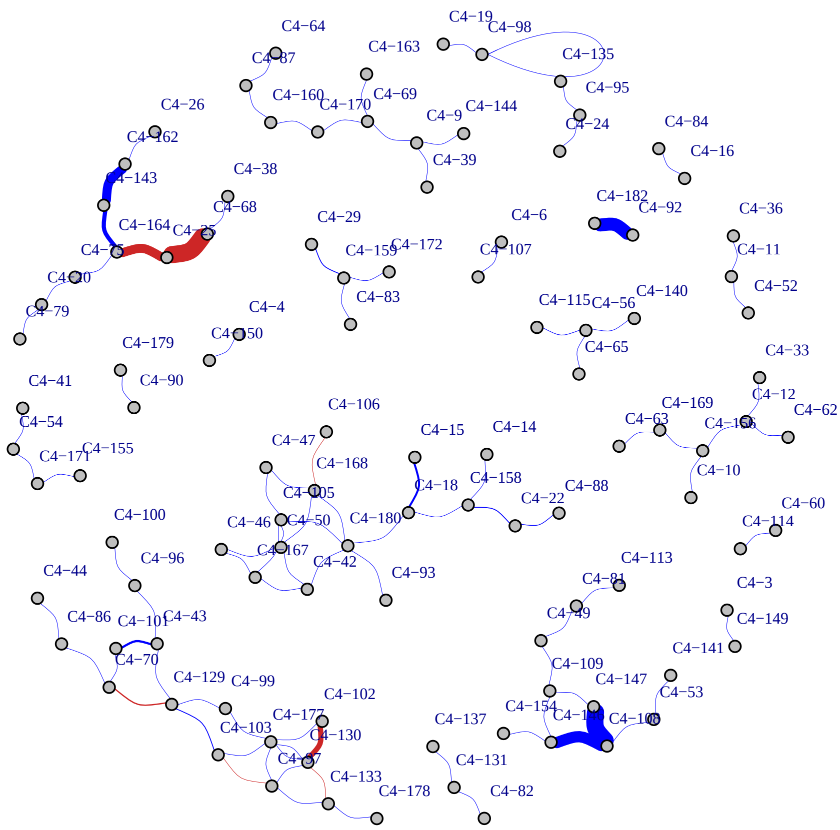

Figure 79. Network graph model for cohort 4 representing the individual level effects from the "influence" linear regression model for the positive affect survey response. Values that are in red are negative influence, values in blue are positive interactions.

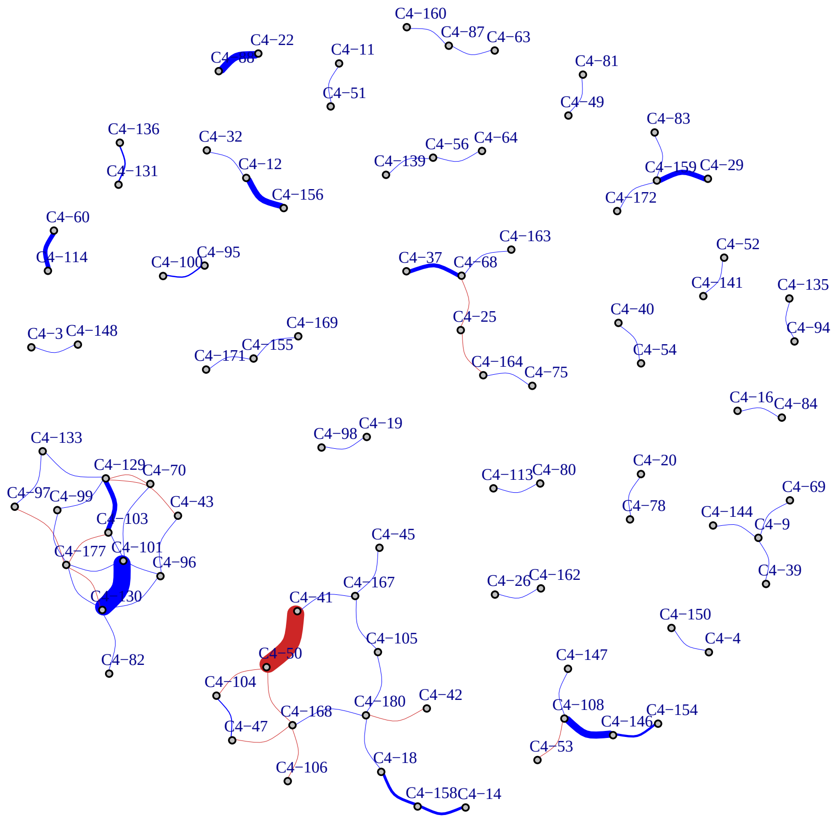

Figure 80. Network graph model for cohort 4 representing the individual level effects from the "influence" linear regression model for the alcohol survey response. Values that are in red are negative influence, values in blue are positive interactions.

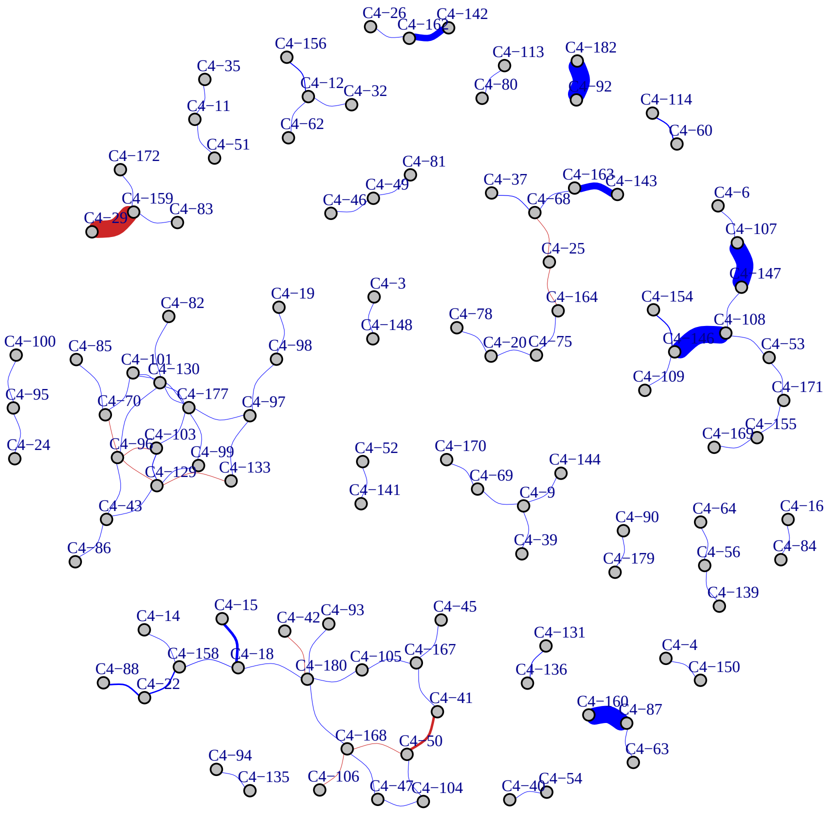

Figure 81. Network graph model for cohort 4 representing the individual level effects from the "influence" linear regression model for the sleep survey response. Values that are in red are negative influence, values in blue are positive interactions.